# COMPLEMENTARITY MODELLING OF INVESTMENT IN ELECTRICITY GENERATION CAPACITY

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MANAGEMENT SCIENCE

IN THE UNIVERSITY OF CANTERBURY

BY

P.R. JACKSON

UNIVERSITY OF CANTERBURY

2015

# ACKNOWLEDGEMENTS

I would like to thank Professor Grant Read and Dr Shane Dye for the magnificent support and mentoring they offered throughout the study process. Their knowledge and ability are beyond doubt. They allowed me the freedom, as a mature student, to independently explore a topic that I found, and still find, fascinating. Naturally, the responsibility for errors and omissions remains mine. I particularly wish to note that their support was forthcoming and generous despite the trying circumstances surrounding the Christchurch Earthquakes of 2010/2011. The earthquakes significantly impacted on the Department of Management, the wider University, and the lives of all Christchurch residents, and to have had the support and assistance I received through this time was testimony to the character of all involved.

My fellow students, Antonio Pinto and Stephen Starkey also deserve significant thanks. Without their input, encouragement and friendship the experience would have suffered. Along with the other members of the Water Market Research Group who provided the opportunity to discuss market related concepts, many thanks are also due to the wider Department of Management for the facilities and support provided.

Finally, thanks to my wife Ellen, and children Frederick and George, who have had to endure my "distracted" ways in recent years. They are a source of joy in my life and consistently remind me of the hope that education affords us all. Thanks also to my parents who supported this project during extremely trying times for all concerned.

# TABLE OF CONTENTS

# TABLE OF FIGURES

# ABSTRACT

*"The purpose of mathematical programming is insight not numbers"*

Arthur Geoffrion (19876)

Complementarity formulations offer the opportunity to thoroughly investigate and clarify the problem of investment in electricity generation capacity. While complementarity has been traditionally used in the sphere of imperfect competition, we demonstrate it also can play a fundamental role in perfectly competitive situations. We demonstrate that our approach offers richer understanding than the traditional linear programming approach. We attempt no judgement as to the practical benefit of our approach, as the benefits themselves depend on the data of the specific counter-factual used. The disadvantages of our approach are somewhat clearer. We acknowledge that the formulations that result are computationally challenging and that the various standard solution methods available for complementarity problems may not actually represent efficient solution methods for such problems. Nevertheless, we adopt the complementarity framework, as ours is a purely theoretical thesis, designed to explore the potential of a unique solution approach to an oft-solved problem. Complementarity theory provides an environment for developing theoretical formulations that, in many cases, resolve directly from an optimisation problem, but it is also free to include other conditions where required.

After a brief introduction and literature review, Chapter 1 considers the finer detail of traditional solution approaches, including the use of screening curves, linear programming, and a related complementarity problem. Screening curves were traditionally used in the times of central planning to describe the optimal trade-off between technologies in terms of utilisation, and with the addition of a Load Duration Curve, the optimal capacity of each technology. We explore various representations of the LDC and discuss how these interact with investment conditions and how market clearing procedures are viewed in their context. We show, with the use of an example, that LP approaches are incapable of accurately defining key system performance measures such as the Loss of Load Probability, without either "guidance" from the modeller or through the use of a significant number of load classes or slices. Furthermore, we show that supposedly perfectly competitive models produce prices that are either inconsistent with the perfect competition they are predicated on, or inconsistent with the optimal capacity suggested by the model. Our investigation identifies the reason for this deficiency.

Optimal trade-offs have a useful theoretical function, but they also emphasise the nature of the technological choice, and ultimately when the trade-off is with a notional shortage technology, they describe the nature of the total capacity choice. But the screening curve approach quickly succumbs to complexity, and is often replaced by optimisation in the form of linear programming, in which a significant number of constraints could be more easily expressed. Nevertheless, the screening curve concept has some conceptual advantages that we can integrate into the analysis, namely the determination of utilisation levels corresponding to optimal technological trade-offs. Knowing that the traditional LP approach does not accurately reflect the relative timing of investment and operation

decisions, or produce solutions that are independent of the LDC definition, we consider the integration of screening curve logic. By way of resolving the downsides of the LP approach, we develop a complementarity formulation that combines the LP solution with the logic of screening curves to derive a problem representation that enables an accurate and consistent solution to the simple problem. In doing so, we make clear that screening curves, per se, are not the motivation, but the vehicle for determining an optimal representation of the system.

Complicating the investor's decision processes are several technological issues, the relevance of which might vary from market to market, but should be considered. Chapter 3 describes a non-exhaustive range of typical problem extensions that would challenge screening curve analysis and how our basic approach can be adapted to include these. This is important for several reasons. Firstly, it is important we demonstrate the overall extensibility of the approach. Secondly, each extension involves discussion of both the extension itself, which in many cases is represented can be accommodated by additional constraints or altered objectives in the underlying optimisation problem, but also the way in which these extensions impact on the definition of the optimal system representation. In deriving the optimal system representation, we develop duality based pseudo-screening curves to describe optimal trade-offs in particular situations or scenarios. By way of example, we consider the relatively standard fare of cost structure generalisation, capacity inflexibility, energy limits and storage, and finally the formulation and interpretation of configurable technologies as non-linear notional technologies.

In Chapter 4 we refocus on load. We consider demand response in two forms: the short-term demand response that typically requires investment and can be written as a technology, and the wider type of demand response that comes as a result of adjusting consumption patterns and substitutions. By nature, the latter response is based on longer term considerations and, in the spirit of the investment problem, we develop an approach to including this response in a fashion that excludes this form of demand response from behaving as a marginal technology in the electricity market. We then consider reliability using an endogenous augmented LDC formulation. Finally, we present a formulation and investment analysis of intermittent generation based on a chronological load and generation pattern. This case requires the introduction of an additional level of dynamic LDC generation, and the maintenance of a dynamic mapping between the LDC and the chronological load pattern.

No discussion of investment is complete without consideration of risk and uncertainty, and it is therefore important to demonstrate how this can be addressed in our formulation, and what the consequences of risk are. We begin by properly defining these terms before expanding the formulation to consider how risk could be implemented. Our over-arching approach is to develop the framework in accordance with the principles of Ralph & Smeers (2011), including contract market, whose clearance defines the market price of risk endogenously. We distinguish between the perspective of portfolio optimisation based on investor preferences and the perspective of risk constraints using an objective function that combines the expected profits of the firm with a CVaR measure. Uncertainty is presented as a distinct concept. Our presentation focuses on various conjectures and the implications they have for the optimal investment condition.

Throughout the thesis, we use the structure of complementarity models as this is both convenient and the basis of much prior research into similar questions. However, complementarity

solution methods per se are not the focus of this research, and we refer the reader to those texts listed for a detailed explanation of the theoretical properties and solution methods associated with complementarity problems. We felt it important to be able to describe the problem within a single framework rather than an ad hoc collection of algorithms, and aim to show that, even though decomposing the problem and using algorithms may be more effective than standard complementarity solvers, complementarity formulations can be implemented for a wide variety of purposes.

# MOTIVATION

Restructuring has seen centrally planned industries become competitive industries with operation of generation assets on a commercial basis, albeit with government ownership remaining in some cases. Investment incentives are forefront in the minds of investors, and those who steward the power system, for it is typically no longer publicly funded. From the widest perspective, the goal of this research is to investigate these incentives. To that end there has been much work conducted. In the post-restructured era, understanding strategic interactions, risk and contracting have become the top priority for practitioners in this field.

By far the most common strategic interaction featured in the literature involves spot market gaming. Absent an over-arching discipline, these games necessarily reflect a short term and opportunistic set of incentives. With few exceptions, these models cannot be implemented in a purely optimisation framework although in some cases optimisation approaches have been identified (Chattopadhyay 2004). More advanced studies have integrated investment, and in doing so have added a significant level of complexity (Murphy & Smeers 2005). These problems typically require complementarity formulations, and depending on the assumptions made MPEC/EPEC formulations (Ralph & Smeers 2006). However, the way these models deal with entry is truncated in the sense that they generally involve consideration of a fixed number of participants without investigating or specifying the basis for the maintenance of market power amongst the firms modelled at the expense of other entrants.

There have also been many empirical studies of electricity market performance. These too have largely focussed on short term incentives in the spot market. Typically these studies have attempted to fit the behaviour of firms to a particular gaming paradigm, for example Cournot gaming, or supply function equilibrium (Wolfram 1999), with the resulting output being the parameters that best account for the observed behaviour. One such example from was the Wolak report (Wolak 2009), into the performance of the New Zealand Electricity Market. Setting aside issues regarding the modelling of hydro generation, the Wolak report found that firms in the NZ electricity market possessed significant market power. This conclusion was reached using an ex-post measure of market power, implying that while the market power existed, it had not been fully utilised (Evans, Hogan & Jackson 2012). This form of modelling is also incomplete, as the lingering challenge arising from the study was understanding why firms with market power had not actually exercised it.

One possible reason is that much of the actual analysis of market behaviour, and the most theoretical gaming analysis presented ignores the true structure of the electricity market and underestimates the most important strategic aspects of the market. A wider perspective of the sector is presented in Figure 1. The number of potential strategic interactions is significant, and far too many to consider in a single thesis. Current research is heavily oriented toward inter-firm strategy, but there are clearly other "games" worthy of consideration, that could explain aspects of electricity market behaviour that current theoretical and empirical studies do not.

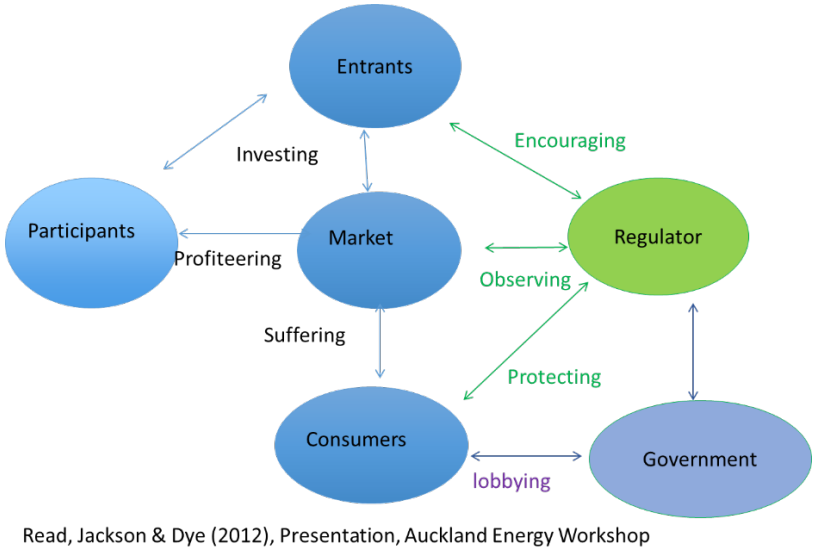



**Figure 1: Typical Electricity Market Structure and Interactions**

Given the number of interactions is large, the immediate goal of this research is not even to explore these aspects directly, but to develop a framework suitable for doing so. A priori it seems reasonable to assume that the same broad modelling challenges would apply throughout Figure 1, and as inter-firm interactions require more advanced techniques than pure optimisation, so it seems will this framework. In addition, some of the most recent advancements in the study of risk such as Ralph & Smeers (2015) involve the development of stochastic endogenous equilibria, supported by multiple participants. Again, pure optimisation will not suffice in such an environment. Accordingly, complementarity theory is the basis for the framework we develop, even if on occasion optimisation would suffice.

Many of the interactions above exist whether or not the spot market is competitive. Given the relatively large amount of research into spot market gaming, we elect to begin by considering interactions in a perfectly competitive spot market. The first step was to develop a framework that would accommodate the basic features that a researcher might want in a model of energy markets:

- Consistent solutions
- Energy and Capacity Limits
- Demand Response
- Configurable technologies
- Endogenous intermittent generation
- Risk measures that align with economic theory

These requirements necessarily involve detailed application of complementarity methods. In developing an understanding of those approaches, some deficiencies in conventional modelling approaches are identified, and opportunities for complementarity theory to be applied to the investment problem even under perfect competition were recognised. A number of other approaches are used throughout the thesis. These include:

- Optimisation models that were used to derive KKT conditions that can be applied in the complementarity framework

- Optimisation models that were used to derive KKT conditions for ranking algorithms or finding minimum next steps
- Optimisation models that were used to derive KKT conditions that define CVaR.
- Screening curve models which are discussed as they represent a different approach that (notably) produces a superior solution to conventional optimisation approaches in simple cases.
- Financial Options, which are discussed analogously with capacity values and therefore can be used along with the mathematics of screening curves to precisely define utilisation levels in scenarios or sub-periods.
- Algorithmic approaches, which are used to guide solutions to, or away from, values that are inappropriate

The final reason for presenting the thesis in complementarity form is that it enables a consistent representation of all of the above approaches in a single model.

# 1 INVESTMENT FUNDAMENTALS

## *1.1 Introduction*

We begin by discussing the motivation for the thesis and the choice. The thesis begins by discussing some of the features of electricity markets and the evolution of issues that have led to the use of various analysis techniques. Naturally this topic spans the entire literature and so our discussion is modest, but it does include a discussion of the use of screening curves, optimisation and complementarity theory as each have been applied to various aspects some of the investment problem. A number of approaches have been suggested and we discuss these briefly before moving to a detailed discussion of screening curve analysis, with a particular focus on the suitability of the many assumptions attached to this analysis. For example, in its standard incarnation, screening curve analysis does not address network constraints, losses, the possibility of economies of scale in generation, or more subtle issues such as the continuity of investment. The appropriateness or otherwise of these assumptions are discussed in the context of perfect competition.

While screening curve analysis is intuitively successful, it has been largely abandoned in the literature as the need for additional complexity has driven analysis beyond its capabilities and has typically been supplanted with optimisation based techniques. Being a relatively vintage approach compared to optimisation, screening curve analysis appears dominated, as its ability to practically address the types of complexity present in electricity markets decreases as complexity increases, whereas the introduction of increasingly sophisticated and data-intensive models has supported increased adoption of optimisation-based techniques to solve complex investment and capacity planning problems. Nevertheless, the original screening curve analysis remains a useful conceptual and analytical tool and we introduce it with a view to capturing the relative strengths of the analysis in an alternative use of complementarity theory, in Chapter 2.

In Section 1.5.3, we discuss the relationships between social welfare optimisation, pareto efficiency, and competitive markets along with the Fundamental Theorems of Welfare Economics. These relationship are necessary to explain which modelled outcomes are, and are not equilibria or pareto-optimal. With these in mind we present a two stage and single stage optimisation formulation in of the investment and generation problem in general terms. In section 1.6, we introduce some of the load representation options available when those general optimisations require specialisation. Optimisation models require some representation of load, and that representation should be accurate and fit for the purpose of the model intended. In this chapter, we only consider LDC based load representations, leaving the implications of chronological load patterns to later chapters. The method by which the LDC is represented and formulated determines the nature and number of products and prices that form the output of the investment model.

We discuss the applicability of these approaches in the context of conventional optimisation formulations. Most importantly, we review the role of the LDC representation in restricting generation functions in conventional formulations. By way of example in Section 1.6.2, we show this either leads to pricing that cannot be supported by the spot market clearing process, or where the spot market

clearing process is assessed against prescribed capacity, we show that the capacity will not recover costs. The case where the LDC is piecewise constant is a special case and we address that also in Section 1.6.3. We find some comparative advantages in screening curve analysis that motivate further research in Chapter 2, in which we introduce complementarity models that are distinct from prior applications of complementarity theory.

## *1.2 Electricity Market Features*

Electricity markets have several properties that, at least partially, inhibit the efficient operation of the market (Botterud & Doorman, 2008). On the demand side, load is highly inelastic in the short term, both because of the lack of alternative energy sources available and the lack of ability to respond to, or even be aware of, pricing signals from the spot market. On the supply side there are a myriad of complicating factors that do not exist, or are not as prevalent, in other commodity markets. Some of the more notable distinctions between electricity markets and standard commodity markets are:

- Electricity is non-storable
- Electrical current cannot be directed or traced and will take the path of least resistance
- Transmission (delivery) is dependent on line capacities, and there are losses incurred in the transmission process
- Electricity production techniques are not homogeneous and cost structures differ significantly between different generation methods.
- Generation plants are involved in complex scheduling arrangements and may not be flexible in the short term.
- Generation equipment participates in several markets at the same time, supplying power, reserve and other ancillary services.
- Certain types of generation are energy-limited and allocate limited and/or uncertain fuel supplies according to opportunity, rather than financial cost.
- Electricity markets can have acute demand-side flaws as in the short term, at least, price signalling is poor and demand is inelastic.
- Due to the national importance of electricity supply, electricity markets are subject to intense political scrutiny.

Perhaps the most fundamental feature of electricity supply relative to other commodities is that it is non-storable in commercial quantities with current technologies (Bushnell, 2003). The lack of significant storage options effectively requires demand to be met instantaneously, with no significant portion of energy demand being able to be serviced from storage as might occur with other products. Beyond the use of stored electricity, the ability to switch to other fuel sources is limited, particularly at short notice. While improving, the ability to respond to price signals in a timescale appropriate to that in operation in electricity markets is limited, and in many cases the pricing information that would guide such a response is unavailable or cannot be practicably analysed in the frequency necessary to respond to highly volatile prices. Highly volatile spot market prices incentivise demand side

participants to contract significant portions of their demand, and in doing so further reduce the incentive to respond to spot price signals. Contract demand is more elastic, however contract prices are significantly dependent on spot market prices, and for many users their flexibility at the time of contract signing is limited to the choice of provider, and does not extend to alternative energy sources.

With inflexible demand, the task of instantaneously meeting load falls predominantly on the supply or generation side of the market. Non-storability implies that, rather than using a single "most efficient" technology coupled with storage solutions, a range of technologies, each with their own operational characteristics, is required. The immediacy of the requirement to serve load also necessitates the provision of contingency services such as reserve of various types, as there is no possibility of using a buffer stock as might be used to mask delays or malfunctions in the supply chain of other products, and the system is physically vulnerable to the effects of outages.

Although it might already be considered complicated enough, generation to satisfy the immediacy of electricity consumption must be transmitted through an electrical network. In addition to increasing the dimensionality of the problem by adding a spatial dimension to generation and consumption, the network structure limits transfers between locations and results in losses.

The nature of the network and the need to instantaneously satisfy load requirements necessitates significant coordination and led to the development of optimisation based market clearance approaches such as first appeared in Bohn et al (1984). A standard electricity market design involves combining bids from load (demand) where available, and offers from generators (supply) where the bids and offers relate to a particular location and time, with the system constraints by a market operator to produce an optimal dispatch. The market clearance optimisation implicitly maximises a de-facto measure of consumer surplus and in doing so generates pricing that reflects the real-time production and delivery costs of electricity, while satisfying the many constraints that enforce the observance of the physical realities and market features that are modelled (Caramanis, Bohn, & Schweppe, 1982), (Hogan, Read, & Ring, 1996).

Having developed the basic mechanics of market clearance and captured the benefits of the superior coordination that market pricing signals enabled, the focus of researchers switched to broader issues of market design and market structure as the economic implications of the physical characteristics of electricity markets were made clearer by formal optimisation. In particular, governments and regulators in many jurisdictions became concerned with the somewhat juxtaposed problems of market gaming and capacity adequacy, with the latter being of particular concern as the basic competitive market structure does not, by itself, address the "missing money" problem. Naturally, the goal was to ensure electricity markets provided efficient allocative signals for consumption and dynamic signals for investment, both at the aggregate level and in terms of the technological mix. Among others, Green (2002) provides an excellent review of the economic foundations of electricity markets.

One of the predominant structural issues has been the issue of missing money (Stoft, 2002). Missing money refers to income which needs to be recovered, and should be optimally recovered during shortage periods, to support the optimal plant mix. The solution to the problem of missing money is fraught with the competing objective of the need to control pricing, particularly at times when

the market is short and the incentives for gaming are high. In consideration of this problem, and noting the associated risks attached to this portion of revenue, Neuhoff & De Vries (2004) conclude that, in the absence of long term contracting, capacity will be lower and skewed towards less capital intensive technologies.

Faced with the issues above, different regulators and governments went about reform with the end result being a vast number of market configurations. Interventions such as price or offer caps and adjustments to the market structure itself have been introduced to address this and other efficiency issues. Ultimately though, it is often dynamic efficiency, or inducing the optimal level of investment and technology mix that is more important, and in those terms market regulators must consider whether moving towards perfect competition in the spot market or developing a contestable market are more important. In this study, we focus on a contestable market, where entry and exit, are uninhibited.

In many markets the question of capacity adequacy did not come into focus for a number of years as many previously centrally planned systems had excessive capacity levels, but eventually growth and retirement eroded the overhang of capacity, and the dynamic issues of market design came into focus (Botterud, Ilic, & Wangensteen, 2005). A variety of market structures have been created to promote capacity adequacy. The large number of configurations possible is partly a result of many, and varied, attempts to address several distinguishing and complicating features in electricity markets. The specific design of capacity markets has been argued at great length. Each implementation was designed in accordance with the particular philosophy and concerns of the body charged with implementing a market reform. Stewart (2007) includes a market structure taxonomy that is useful when describing many of these possible market structures.

These market designs, which attempt to simultaneously address revenue associated with shortage while promoting capacity adequacy, are a source of significant difference between individual electricity markets. A number of proponents, have weighed in on this fundamental structural issue, particularly in the aftermath of the Californian power market failure (Cramton & Stoft, 2005). Among the options suggested were:

- Energy only – prices are expected to spike during these periods to sufficiently high levels to provide revenue adequacy for peaking technologies.
- Capacity payments – firms receive a capacity payment for having their plant available.
- ICAP systems – firms are incentivised to hold capacity, although not necessarily to offer it.

Summaries of various market structures in operation throughout the world are available from Stridbaek (2005), (Oren, 2005). In both NZ (Layton, 2007) and Australia (AEMO, 2010), the chosen market design calls for the energy price to rise to high levels to cover the capital costs and required return of investment in peaking capacity. In the case of Australia, those levels are somewhat sculpted by both a price cap and a limitation on the length of time in which shortage pricing can occur, but the parameter settings used are derived from precisely the sort of analysis that promotes a balanced trade-off between capital costs and returns for peaking plant. As both local market designs are supported by that

fundamental logic, our study is focussed on single payment markets where capital cost recovery primarily comes from the energy price.

The many features of electricity markets are well known, and much has been written about them. However, the literature does not exist in a vacuum and has been responsive to the various developmental stages of electricity markets throughout the world. Beginning with the design of market clearance mechanisms, and continuing through various phases of development researchers, market regulators and market participants have sought to address the issues of most relevance. To that end complementarity theory has presented itself as an approach with significantly greater flexibility than standard optimisation affords, and this is particularly true when examining issues that require consideration of more than one objective function. Many applications of complementarity theory have focussed on the behaviour of strategic market participants. The analysis of perfectly competitive situations has been, and continues to be, handled by optimisation techniques such as linear programming.

## *1.3 Investment*

### 1.3.1 Discounted Cash Flow

A basic approach to project investment analysis is the maximisation of discounted cash flows. The revenue and expenditure streams associated with a project are forecast and re-valued to a consistent time period (usually present day) using a discount rate. If the net difference is positive the project will be of benefit, whereas if negative the project should not go ahead. Choosing the level of the discount rate is in itself a complex decision and different choices may result in different projects being selected or mothballed (Baumol, 1968).

The inclusion of risk aversion in this type of model can be achieved in a variety of ways, such as modification of the discount rate, or through some re-allocation of scenario weights. Both approaches can be used to penalise investment however the former method is ill-conceived as it concatenates the adjustment for risk and the time value of money by compounding risk along with the discount rate, reflecting a particular, and generally unintended, risk structure. It may be that risk does vary over time but if that is the case, it is preferable to make the necessary adjustments independently, as risk is unlikely to compound exponentially at the risk-free rate of return. To avoid issues and complications with compounding risk adjustments, it is preferable to use a risk free discount rate and then apply additional measures to account for risk (Robichek & Myers, 1966).

*Real Option Theory*

Before discussing real-option theory, we should note Wallace (2010) demonstrates that stochastic programming techniques already implicitly incorporate real options and are therefore not only capable of representing real options, but also identifying them where they previously may have been overlooked. Although multi-stage stochastic programming approaches have always implicitly evaluated real options and potentially identified some options that were previously not well understood, the advent of real option theory did benefit the investment literature by clarifying and specifying the

concept that there may be value purely in the option to invest, if the investor can control the timing of the investment (Dixit, 2012). Under this approach, the investor does not invest when the discounted benefits are merely positive, but waits until the discounted benefits exceed the value of the option to invest.

The dashed line in Figure 2 shows the NPV of the project, which becomes positive when the NPV of project returns exceeds the investment cost, I. The solid line shows the expected NPV of the investment opportunity. While the value in investment is positive beyond I, but there remains the value of the option to invest. As the project becomes increasingly in the money, the additional value accruing from postponement of investment withers until eventually, at V*, the two lines coincide and the value of the project is equal to the value of the option to investment. At this point, there is no value in postponement and the project, with positive NPV, should go ahead. Under traditional NPV analysis this would have occurred earlier, at I, the point at which the NPV of the project became positive. The "real" option is equivalent to an American call option on the right to invest.

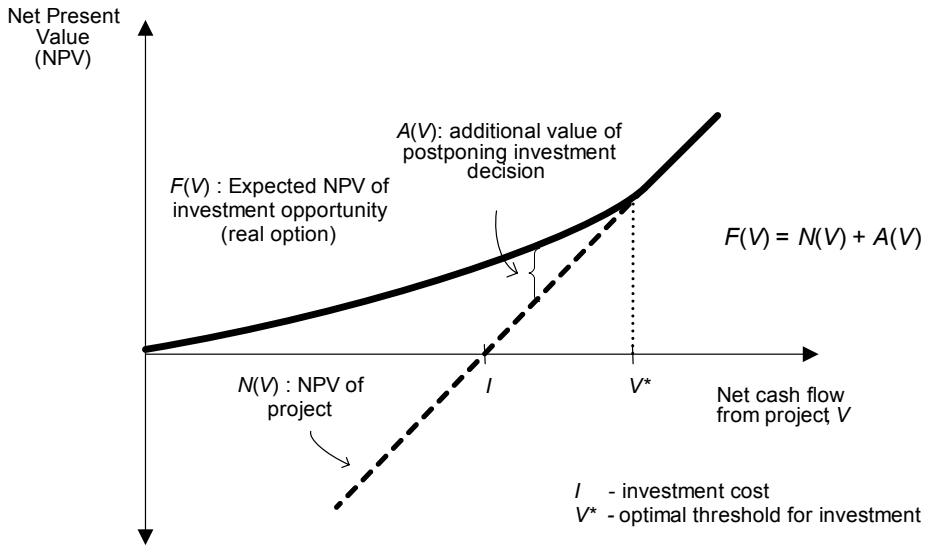



**Figure 2: Real Option Theory**

Competitive and strategic pressure often limits the number of investors that are in a position to choose the timing of their investment (Botterud & Korpås, 2007). Absent entry restrictions, the ability of individual investors to determine entry timing is muted, and in the theoretical limit, zero. While subtle, the assumptions required to support entry restrictions have implications for the spot market that remove the analysis from the perfectly competitive sphere in many cases. For this reason, we do not consider real options in conjunction with investment decision making, however in an advanced setting real options may be useful in consideration of mothballing or reinstatement decisions where the temporary "entry" and "exit" decisions are under the control of a single entity.

### 1.3.2 Wider Perspectives

Majumdar & Chattopadhyay (1999) suggest other issues with the basic DCF approach to investment analysis. The main objections they raise are that there should be consideration given to firm value

maximisation (for direct stockholder benefit) and also to the actual financial constraints that apply, and the implications investment may have on the financial standing or rating of the firm as it may relate to future funding costs (Evans & Guthrie, 2005). Investment schedules may be significantly different under different objectives or pressure of financial constraints. We also note that the public sector investor is potentially less concerned about these sorts of issues, and although we do not address the potential for crowding out in this research it is an interesting aspect of the problem.

Where a firm is listed on a stock exchange, the firm has some additional considerations. The question of whether it should manage risk on behalf of shareholders who are free to strike their own portfolio for that purpose is one such issue. The consequence of risk management is lower expected profits, and potentially duplication, or inefficient risk management where the options available to the firm are different to those of its shareholder. The alternative view recognises that while shareholdings are atomistic and able to be dynamically hedged in theory, in practice they are not, and the transaction costs of managing a portfolio are high and volatility is not attractive for that reason. More importantly, shareholders are not the only stakeholders and in practice relationships with financial institutions, and the covenants they impose in return for advancing funding, require risk management by the firm to ensure basic financial obligations are met.

### 1.3.3 Traditional Planning Methods

One of the first papers on investment in electricity generation systems was Masse & Gibrat (1957). They recognised the distinction between different production technologies, their flexibility, and the differing capital costs of each technology. Traditional electricity planning methods involved minimisation of the cost of supplying electricity, while maintaining a reserve margin in accordance with a chosen standard. To reflect risk aversion, this standard would not be an average year, but a dry or high demand year, depending on the characteristics of the system under analysis. Cazalet, Clark, & Keelin (1978) investigate the optimal setting of this margin. They discuss electricity planning in terms of cost minimisation with respect to the reserve margin. System cost is expressed as the sum of environmental cost, fixed costs, variable costs, and outage costs. All are functions of reserve margin, suggesting that there is an optimal level of reserve margin that minimises the total of these costs.

Turvey & Anderson (1977) provide an overview of investment models in the electricity sector. Following a large amount of research on the optimal configuration and operation of power systems from a centrally planned perspective, Caramanis et al (1982) showed how appropriately set prices could improve signalling relative to standard tariff systems and, in a follow up paper, how spot pricing could guide investment decisions (Caramanis, 1982). From the perspective of actual planning tools, Nakamura (1984) reviewed the then current systems for capacity expansion planning including WASP (Wien Automatic System expansion Program) and the system developed by EPRI, who built a model called the under-over model and incorporated low, medium, and high load growth probabilities. It also incorporated three levels of hydro production and breakdown probabilities, and different reserve levels. In similar fashion, Read, Culy, Halliburton, & Winter (1988) describes the PRISM simulation model that was used for electricity planning in NZ, which represented an advance in the level of detail in operational planning relative to existing planning models by simulating operations on a weekly basis. In an even wider context the US NEMS system represented a comprehensive equilibrium model

extending across the wider energy sector, in which individual components were solved separately in an attempt to provide a consistent analysis of the sector (Murphy & Mudrageda, 1998). Finally, in NZ, Bishop (2007) describes GEM, the generation expansion model that replaced PRISM.

The above approaches were traditional in the sense of their orientation towards planning at the system level. The introduction of new paradigms in ownership and industrial organisation has necessitated new approaches that deviate from the traditional planning approaches and are oriented towards new technologies and new ownership structures. For example, Awerbuch (2006) contrasts traditional least-cost approaches with portfolio based approaches, and focuses on the integration of new renewable technologies such as wind and solar power that must be considered in terms of the portfolio of generating plant that will supply electricity when renewable energy sources are not available.

## *1.4 Screening Curve Analysis*

Screening curve analysis highlights the potential for different technologies, with differing capital and operating costs, to serve loads that occur with different frequencies. Although the analysis may be extended, first and foremost the emphasis in screening curve analysis is on the comparison of costs, and the definition of optimal trade-offs and marginal operating ranges for each technology. It is assumed that load and generation exist at a single node, that technologies are dispatchable, reliable, have effectively unlimited fuel supplies, and that their cost structure may be represented with a linear total cost function, in which the intercept represents the fixed cost of providing capacity and the slope represents the marginal cost of operation. We now consider the implications of the implicit assumptions involved in screening curve analysis.

### 1.4.1 Network Structure

Screening curve analysis assumes all generation and load occurs at a single node. The reasonableness of this assumption rests on whether electricity is being transmitted large distances and whether the network topology results in significant bottlenecks in either energy or ancillary service markets that might restrict the transmission of electricity. The geographically smaller the network, and the closer load centres are to generation nodes, the more justifiable the assumption of a single node becomes. The assumption removes the need to consider complex network structures, but comes at the expense of overlooking the impact of transmission losses, or network constraints. Failing to create a distinction between load and generation locations, and thereby ignoring losses, leads to underestimation of the combined cost of generation and delivery and fails to account for the impact of network constraints, which are potentially significant.

*Network Constraints*

The impact of network constraints on market outcomes is well studied under both competitive and gaming conjectures (Joskow, 2000),(Borenstein, Bushnell, & Stoft, 2000). Where gaming of a network constraint was to occur, the situation is complex, and spot market outcomes would depend on the nature on the market power of the participants as well as the availability and efficacy of spatial hedging arrangements. But even under perfect competition, network constraints distort prices and profit

opportunities in the short run. Price differentials resulting from binding network constraints should ideally provide an investment signal to either investors in generation capacity or transmission capacity. Prospective investors with rational expectations will anticipate that wherever entry is free, either generators will adjust their locational investment strategy or the system operator will adjust network investment decisions to mitigate or eliminate these inefficiencies. Accordingly, the returns they perceive to be available should not include any rent associated with network constraints. There may be specific circumstances that prevent one or other of these adjustment mechanisms from operating, but for the purpose of our analysis we assume the omission of network structure does not introduce significant long-term bias in a competitive market with free entry.

*Transmission Losses*

The transmission of electricity involves losses which vary throughout time and space in a network, with the implication being that what is required to be generated exceeds that which is consumed. Transmission incurs losses that increase quadratically with the amount of electricity being transmitted, and so are greatest at those times when transmission is greatest. Ignoring this issue biases the analysis. While a fully worked model of the network would account for this issue more precisely, we seek a first order approximation to limit the bias in our single node representation. One option is to adjust marginal costs to account for average system losses or, where a relationship between the system loss percentage and load can be established, we could enhance the approach by making the system loss percentage a function of load.

$$\mathrm{MC}_i = \frac{\mathrm{FuelCosts}_i + \mathrm{VariableOperatingCosts}_i}{1 - \mathrm{AverageLoss\%}} \qquad \forall i \qquad (1.1)$$

This reflects the difference between the cost of generating load, and the cost of generating and transmitting to the point of withdrawal, and while not precise, it is not as imprecise as assuming zero losses by using a fuel cost alone. The primary weakness of that approach, or indeed the standard approach of ignoring losses, is that it does not factor the capacity implications of losses. The adjustment above addresses only the generation cost issues, as if generators could produce the electricity eventually lost without using capacity. A preferable treatment involves treating losses directly, so that the extra costs of the requirement for additional capacity as well as generation are both recognised. By way of example, we propose adjusting load proportionally by the same proportion as we adjusted costs in (1.1).

$$L_k = \frac{\mathrm{Load}_k}{1 - \mathrm{AverageLoss\%}} \qquad \forall i \qquad (1.2)$$

That reflects the assumption that absolute losses are a fixed proportion of load. This relationship is only one possible assumption from among many. Whether maximum losses occur at the time of peak load depends on the system but, a priori, and without any further information, the assumption seems reasonable. In the case where a more definitive relationship between load and losses can be established, or the relationship can be specified on a scenario basis such as in terms of hydrological or demand conditions, then the adjustment could obviously be specialised to reflect that information.

### 1.4.2 Cost Structures

Fundamental to screening curve analysis is the comparison of fixed and variable costs.

*Operating Costs*

The marginal cost of a single unit of technology i generating with a 100% utilisation factor is denoted $MC_i$ throughout, and is comprised of two components:

$$MC_i = FuelCosts_i + VariableOperatingCosts_i \qquad \forall i \qquad (1.3)$$

Fuel costs are expressed in as a fixed cost/unit of output, implicitly reflecting the assumption of constant generation efficiency. Variable operating costs are costs other than fuel, such as maintenance costs, that are directly associated with usage and, in this simplest case, proportional to the output of the plant concerned.

*Capital and Fixed Operating Costs*

Fixed operating costs are those that are incurred irrespective of whether, or how much, the plant generates. The sum of capital and fixed operating costs for a single unit of capacity of technology i is expressed as an annuity and is denoted $FC_i$ throughout.

$$FC_i = CapitalCosts_i + FixedOperatingCosts_i \qquad \forall i \qquad (1.4)$$

The capital cost, or installation cost, of a power station is significant, and must be amortised over the life of the plant. The amortised capital cost is dependent on several factors, each of which are individually difficult to determine. The actual build cost of a specific plant may be susceptible to significant variability, as might the cost of maintenance and decommissioning. Even when standard costs are reliably known, or a plant is built as a turnkey development with a guaranteed installation cost, amortising that cost over the lifetime of the plant remains a complex task dependent on the discount rate used in the calculation, and the assessed economic life of the plant.

Determining an appropriate discount rate for a particular project is a complex decision and different choices may result in different investments being selected. Discount rates are often chosen to mimic commercial assessment of projects, where there is a requirement to accurately estimate both the cost of capital and the risks of each project (NZEC, 2007). The Weighted Average Cost Of Capital (WACC) is a commonly used measure of the cost of capital, in which different levels of risk, and therefore rate of return requirements, apply to each capital source, and these are weighted according to the proportion of project finance arising from each capital source. As stated earlier, a superior approach is to separate the time value from of money from the risk involved and assess risk separately. This approach discounts all project cash flow at the same risk free rate before introducing additional adjustments to account for risk and/or uncertainty. To be clear, that is not to say that any financier or funding source would assess their involvement in a project at the risk free rate. The higher the interest rate actually charged for finance, the higher the cash outlays will be, and it is these repayments, not the upfront cost of installation, that represent the cost of the project and must therefore be discounted at the

risk free rate in line with the income stream. Assumptions are also required with respect to the economic life span of the plant, which may be significantly different to its physical lifespan. The economic life of the plant may depend on salvage costs, technological developments throughout the life of the plant and maintenance costs.

In combination, an assessment of these factors enables the estimation of the annuity period, and the fixed cash flow associated with the project, which in turn enables the calculation of an annuity that is equivalent to the discounted costs of providing a reliable unit of capacity for a particular plant type for the lifespan of the plant. Detailed analysis should also consider the implication of interest costs, taxation and depreciation on the project's cash-flows, and may well extend to consider the dynamic impact of any resulting restrictions on cash-flow or the firm's ability to borrow, or any other bank covenants that might come with funding (Majumdar & Chattopadhyay, 1999). In that respect, neither screening curve analysis or our equilibrium analysis deals with the degree of specificity required for a particular investment decision by a particular entity. Instead, standardised amortised capital costs are assumed, and no attempt to model firm balance sheets and banking sector objectives is made. This seemingly limits the analysis however participants seeking to invest may have little, if any, more information on which to base their assessments of market equilibrium. That equilibrium is ultimately determined by their collective investment decisions, whether or not those decisions are based on complete, or partial information, so from an equilibrium perspective, the lack of accounting for the financial structure of current and future investors in generation capacity may not be significant.

*Continuous vs. Discrete Investment*

It is implicitly assumed that investment in any technology is continuously available at a constant cost per unit of capacity installed. However, capacity is unlikely to be available on a continuous basis as most technologies are available in discrete installation sizes. This may be a physical issue, related to technological requirements or the installation location, for example. Alternatively, it may arise as a result of construction and installation economies, which suggest that the basic unit of installation for each technology is determined by the most efficient configuration of that technology.

We are interested in investment at the technological level so that the assumption of continuous investment is justifiable in markets that are large enough in comparison to individual investment opportunities. In smaller markets, the requirement to choose between discrete installation capacities may be significant, as plant sizes prescribed by a continuous model may not be feasible. As the market size grows relative to individual plant sizes the bias or error introduced by this assumption becomes less significant, as any amount of capacity in a particular technology can be approximated by some combination of available discrete unit sizes with increasing accuracy.

Nevertheless, even in larger markets there are complications that could be material and should be considered. As discussed, network constraints may fragment a large market into portions in which the issue of discrete capacity choices becomes significant once again. There is also a bias in the analysis against those technologies that are scalable, and whose flexibility in installation terms goes unrecognised. Similarly, there is a bias towards those technologies that require large installations, whose inflexibility goes unrecognised. We also note that there is also some additional benefit from the perspective of risk management in being able to follow load growth more closely by scaling a flexible

technology as load uncertainty reveals itself, rather than more engage in sporadic investment in lumpy technologies.

In the worst case, the assumption of continuous investment could lead to feasible solutions that are infeasible. Where plant size restrictions do not reflect infeasibilities as much as economies of scale in construction, solutions may recommend installations of a size for which the installation and/or operating costs have been incorrectly assessed. While discrete investment opportunities have been analysed for some time (Levin, Tishler, & Zahavi, 1985), we prefer to rely on the characteristic nature of perfect competition and proceed on the assumption the market is large enough relative to investment units to justify the assumption of continuity in capacity.

### 1.4.3 Optimal Trade-Offs

The total cost of operation at a given level of utilisation for a particular technology is defined by the sum of the two cost components, as shown in (1.5). This pro-rata total cost function expresses the total cost of building and operating a 1 MW plant, as a function of either the number of hours a year the plant operates, or alternatively as the fraction of time, $u$, for which a plant will operate, which we refer to as the plant's utilisation factor, or level. The associated marginal cost is scaled according to the definition of utilisation chosen.

$$TotalCost_i = \mathrm{FC_i} + \mathrm{MC_i} u \qquad \forall i \qquad (1.5)$$

$$TotalCost_j = \mathrm{FC_j} + \mathrm{MC_j} u \qquad \forall j \qquad (1.6)$$

Considering the fixed and variable cost of two thermal technologies with constant efficiency, we can calculate the utilisation level at which one will supplant the other as most efficient technology for a given operating role, or utilisation factor. At its most simple, this trade-off is independent of the load duration curve (LDC), and depends only on the relative fixed and variable costs of each available technology, as described below:

$$u_{i,j} = \frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} \qquad \forall i, j \qquad (1.7)$$

The level of utilisation at which one technology becomes preferable to another is defined mathematically in equation (1.7), and geometrically in Figure 3, by the intersection of the respective cost functions. Among the set of all such pairwise technological comparisons, the most relevant trade-offs and utilisation levels are those defined by the intersection of total cost curves on the lower envelope of the screening curve diagram, as shown in Figure 3.

For each operational role, the lower envelope of total cost functions defines the least cost method of production, and identifies the technologies best suited to that role. Capital-intensive technologies with lower fuel costs are more likely to be appropriate for the satisfaction of loads that occur with high frequency, while less capital-intensive technologies with higher fuel costs are more likely to be appropriate for satisfying load that occurs with lower frequency. The screening curve framework also provides a graphical interpretation of the situation for technologies that are dominated at all meaningful utilisation levels. Graphically, these technologies play no part in the lower envelope.

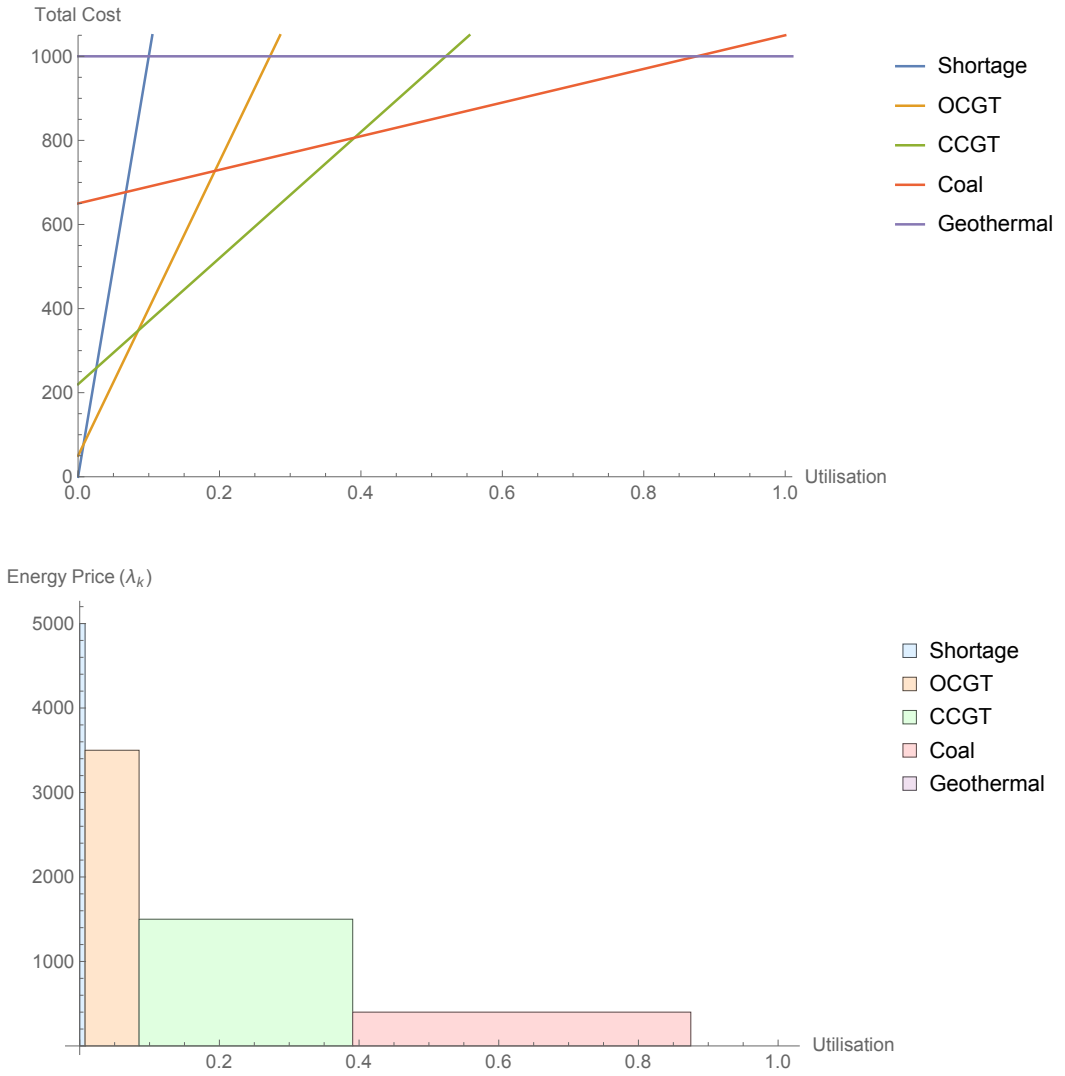



**Figure 3: Screening Curve & Price Duration Curve**

As shown in Figure 3, if we additionally assume the market is perfectly competitive, the screening curve analysis also defines the underlying price duration curve (PDC), which represents a cumulative distribution of prices. For the utilisation range in which a particular technology defines a segment of the lower envelope of cost functions, then that technology is marginal and sets the system price, which in the case of perfect competition is equal to its marginal cost. Because the utilisation range for which each technology is marginal is a function of the full set of technological cost structures and not load, the price duration curve, which expresses both system prices and price durations does not depend on the specification of load in this simplified case.

### 1.4.4 Shortage Costs and Frequency

Screening curve analysis allows for shortage to be viewed as if it were a technology with an associated marginal cost. Although the cost of shortage, or the value of lost load (VOLL), is very difficult to define, once defined, this approach allows a straightforward representation of the critical economic

trade-off between the cost of shortages and the cost of providing more capacity. We proceed, assuming a single and known shortage cost (VOLL), and that the market price is free to rise to this level. We treat VOLL as the marginal cost of a notional shortage technology, which has a fixed cost of zero. Just as we can determine the optimal trade-off between different generation technologies by comparing the cost functions of each technology, we may do the same when treating shortage as a notional technology. As we consider lower and lower utilisation levels, eventually a peaking technology will represent the least cost production method for satisfying very infrequent load requirements. The installation and operating costs of the final peaking technology must be weighed against the benefits it provides, and the frequency with which those benefits are provided. Eventually, as capacity of the peaking technology increases and shortage events become rarer, shortage events will become so rare they are no longer worth avoiding with further investment in generation capacity.

Although in practice the setting of shortage prices is significantly more complex, the shortage cost, VOLL, also theoretically translates to the PDC, and defines the system price that, when offered for a period of time commensurate with the shortage frequency, provides capital recovery for the peaker of last resort. The determination of the level at which shortage costs should be set is beyond the scope of this study but we make the following comments.

Shortages are unable to be targeted, or allocated, to individual customers, and instead affect a range of consumers who will each place different valuations on the interruption of supply. We commence with a single marginal cost of shortage but more elaborate shortage cost functions could be employed, using techniques described in Appendix 7.5. But irrespective of the form of the shortage cost function, in the absence of a demand response scheme, economic rationing cannot be implemented and the expression of the cost of shortage must reflect the unpredictability of shortage events and the inability to ration them.. For that reason, a complete assessment of actual shortage costs, which could be used to strike an overall economic trade-off, is of limited use.

Left to itself, system reliability is generally a public good. However, in cases where it is deemed worthwhile, as assessed by comparison of cost and potential savings in the wholesale or contract pricing, some consumers may opt to install technology to enable load response. Demand response technologies reduce allocative inefficiencies in times of shortage and, more generally, at any time where the system price exceeds the consumer's valuation of the energy. By broadening the concept of shortage and shortage costs to include more general concepts of demand side response it is possible to incorporate a more sophisticated view of shortage by defining additional notional technologies, with non-zero fixed costs, that each represent a particular demand side response opportunity and are parameterised by a price, capacity and perhaps a limit on the total energy available for response within a defined period of time. These ideas are developed in Chapter 3.

### 1.4.5 Optimal Cost Recovery & Asset Valuation

We can view the screening curve diagram and the associated equilibrium PDC from either the perspective of cost recovery, or asset valuation.

*Cost Recovery*

As shown in Figure 3, the lower envelope of the screening curve diagram describes the least cost of production subject to meeting load occurring with a given utilisation level. This least cost may be arrived at in two ways:

- The cost of supplying (an incremental MW of) that load is given by the cost of building and running the marginal technology used to supply it. Diagrammatically this involves moving along the cost curve of the relevant marginal technology.
- The financial cost, which must be equal to the cost of actual supply in this equilibrium, may be found by tracing around the lower envelope, starting from the origin. This represents the marginal cost of buying in this power from other sources, some at marginal production cost, some from technology 1, some from technology 2 etc. Equivalently, this represents the marginal value of 1 MW of that kind of capacity to the system.

Accordingly, for an optimally planned system, and every utilisation level, the following are equal:

- The plant level long run marginal cost (LRMC) of building and operating the specific optimal plant (fixed plus variable) best suited to meet load occurring with that frequency in the long run (as calculated by the first approach)
- The system level short run marginal cost (SRMC) of operating the optimal plant mix (variable only) required to meet load occurring with that frequency in the short run (as calculated by the second approach)

If the system level SRMC and the plant level LRMC were to differ, then there would be a clear preference for switching to a cheaper procurement method, whether that was simply purchasing electricity from the system, or constructing a plant. In terms of the Fundamental Theorems of Welfare Economics, which we discuss in detail in Section 1.5.3, any solution for which this system level SRMC and the plant level LRMC differed would not represent a Pareto-optimal solution, and therefore could not be considered a perfectly competitive equilibrium. Furthermore, in a perfectly competitive economy, energy prices are set by the highest short run marginal cost for any of the stations producing at the time so that by the equivalence of the two cost measurements in equilibrium, the costs of investing in an optimal plant mix (both fixed and variable), are exactly recovered. This is reflected in the equilibrium price distribution, which provides exact recovery of all fixed and variable costs for all plant types, in a fashion consistent with the choice of VOLL/shortage frequency. This is what should be expected in a risk-neutral perfectly competitive market with free entry. As will be discussed in Chapter 5, when investors are risk averse or have an aversion to uncertainty, entry requires more than simple cost-recovery.

*Plant Valuation*

As an alternative to the cost recovery perspective, we may consider a financial or valuation based perspective. From this standpoint, thermal plants represent a strip of physical options to "call" on that plant when the value of its production, in terms of either avoiding shortage or the need to generate from more expensive plant, exceeds its own marginal production cost. Thus, the value of thermal plant can be defined by the value of a strip of call options with the strike price set at the short-run marginal cost (SRMC) for that plant. Where there is also variability in the cost structure of the plant, then rather than

considering a strip of call options, the plant could be viewed as a strip of spark spread options. In a perfectly competitive market, the plant earns no contribution to fixed and capital costs when it is the marginal generator as it will be offering to the market and consequently defining, and receiving, a market price that reflects only its marginal cost. Instead, plant profitability (and cost recovery) is based on those periods when the plant is infra-marginal, at higher load levels, when other technologies are called into service, resulting in higher prices. At these times, we would describe the option as being "in the money".

Based on the proportion of time each particular technology is marginal, we can calculate the time weighted average price (TWAP) received by the plant when it is profitable. The operating value of the plant, per MW of capacity, is given by the accumulated profit of the plant, which is the area of the price distribution above the option strike price. In equilibrium, this area will match the capital and fixed cost of providing capacity of this type.

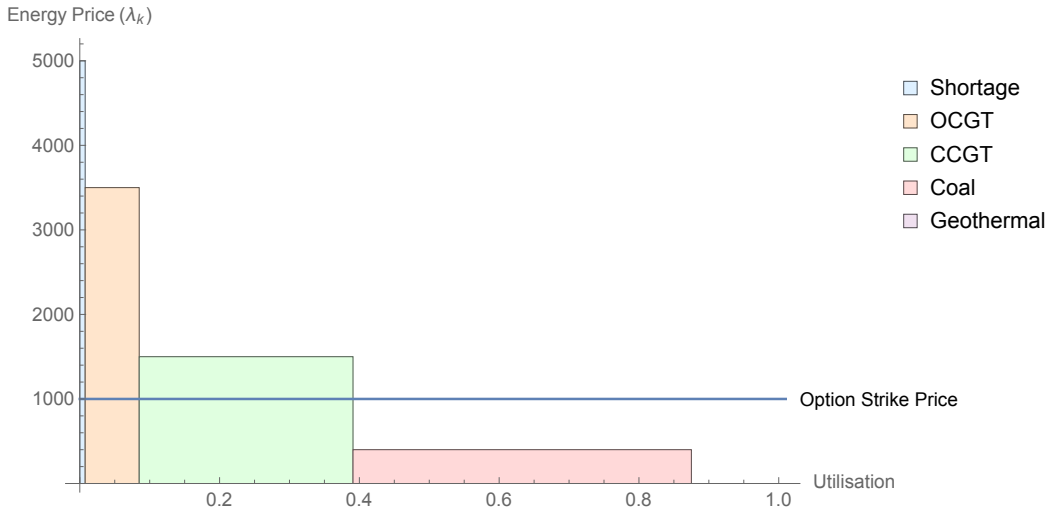



**Figure 4: Call Option Valuation of Thermal Plant**

Figure 4 shows the valuation of a prospective new technology with SRMC=1000. By construction, this particular distribution of returns is entirely deterministic, although very little complication of the model will the make the PDC dependent on the LDC which itself is the source of significant variability. In any case, this concept, while theoretically redundant in this simple case, is useful in more realistic formulations and enables the owner of this plant to approximate the plant as a strip of call options set against a backdrop of whatever scenarios are of interest. Such a representation is just an approximation though, as the valuation of thermal technologies as call-options is complicated by the fact that physical generation assets do not behave exactly like financial contracts. In practice, there are a number of non-convexities and constraints, such as start-up costs or ramp rate restrictions, that inhibit the free exercise of the physical option relative to that of the financial option. Nevertheless, in our framework, absent chronological development as it currently is, the analogy is reasonable and proves very useful as we impute the value of capacity in specific sub-periods, or under certain scenarios.

## *1.5 Optimal Plant Mix*

### 1.5.1 Load Duration Curves (LDC's)

The preceding section described how, under certain circumstances, we can define the equilibrium utilisation ranges for each plant type, without reference to load. To extend the analysis and prescribe a plant mix requires a representation of load. The most common of such representations is the Load Duration Curve (LDC). Beginning with a chronological load pattern, periods can be re-ordered to form an LDC, representing the cumulative distribution of load, with each utilisation level describing the fraction of time for which load is at least equal to the corresponding load level. In the case of our investment model, the relevant time period is a year. From the perspective of investors, the LDC is not simply historical load levels as measured, for example, in half hourly, or five minute scheduling intervals in a real market. Instead, it is the convolution of the probability distribution of load in each of those time periods.

While the LDC representation is common, and analytically helpful, we must be mindful of the inherent weaknesses of this form of load definition. The underlying structure of load variation is typically the cumulative result of daily, seasonal and annual patterns. Construction of an LDC destroys information about the chronological nature of variations across many different time scales, each of which have different drivers, and implications for investors and technological operations. The significance of the bias introduced by this approach depends on the nature of the system under consideration. At one extreme, if the overall load pattern is repeated precisely on a daily basis, then to satisfy load requirements each and every day, each plant would have to start-up, perform its role, potentially for a very short timeframe, and shut down again. In practice this would be untenable. For example, a true peaker of last resort for perhaps only a few minutes every day to satisfy the load pattern. Aside from any possible physical constraints further constraining operations, it may also be the case that start up and shut down costs alone make certain technologies, which appear to have a role, uneconomic in reality. At the other extreme, if the overall load pattern was representative of a twenty or thirty year period with variation dominated by multi-annual cycles, then there would be fewer operational issues but instead of load reliably reaching very high, shortage inducing, levels for a small portion of time on a regular basis, investors might only earn high prices during a single crisis of extended duration occurring every ten or twenty years. When viewed from the perspective of a risk-averse investor, these polar investment propositions are clearly different, requiring some careful consideration when determining the most significant features to include in an investment model.

The extent of the bias will be less than that suggested by the polar examples above, although in a particular market the implications of these issues should be considered. To avoid initial complications and maintain consistency with a majority of the literature, we adopt an LDC based approach before extending our framework to incorporate seasonal patterns in Section 2.3.6, and the integration of chronological representations in Section 4.4 in an attempt to minimise the exposure of the analysis to the potential bias introduced by the definition of the LDC.

### 1.5.2 Graphical Optimal Plant Mix

As shown in Figure 5, in the simple case we have presented, the MW capacity of each plant type that is required to meet the range of load requirements can be shown graphically. The optimal installed capacity for each technology is determined by the optimal utilisation range for that technology and can be read off the left hand axis of the LDC. The optimal installed capacity of each technology type will be affected by changes in the LDC arising from growth and/or load shifting however, in this simple case, changes in the LDC do not alter the equilibrium utilisation of each technology, as this is independent of the LDC and depends only on the economic trade-off of fixed and variable costs for each technology.

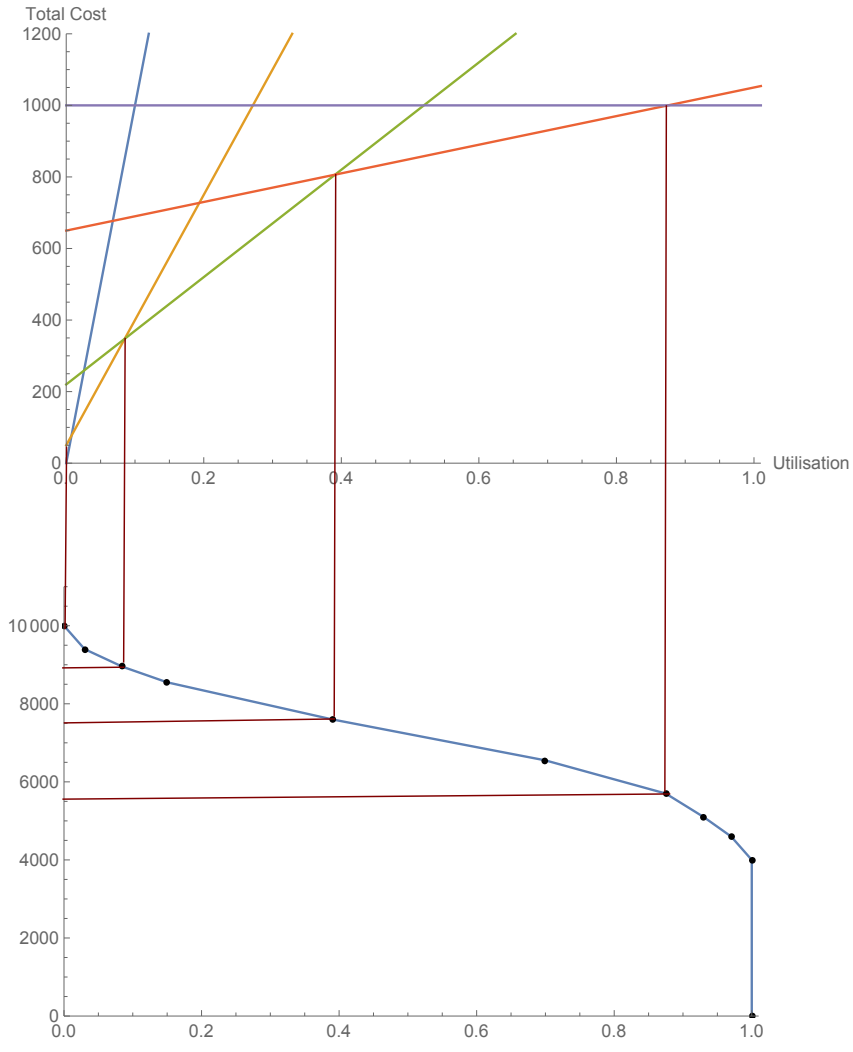



**Figure 5: Graphical Derivation of Optimal Plant Mix**

As shown, and by construction, the graphical approach determines the relevant optimal technological trade-offs and then involves reading the LDC to determine the capacity of each technology that is required. As we advance to a general optimisation format, the LDC becomes an exogenous part of the optimisation that, in general, will not be defined in terms of the points corresponding to the optimal

trade-off between technologies. Figure 5 shows some such points on the LDC that do not correspond to optimal technological trade-offs.

### 1.5.3 General Formulations of Optimal Investment

The graphical solution method shown in Section 1.5.2 is intuitive and remains a useful tool for visualising solution but, as assumptions are relaxed, the model is generalised, and additional complexity is incorporated, the solution is difficult to depict graphically and the determination of an optimal plant mix requires a formal model. We proceed with the following set of basic assumptions, which allow us to focus on the critical issues:

- All trade is assumed to take place at a single node,
- There are no economies of scale in generation,
- Investment and generation are continuous, without non-zero minimums,
- The spot market is assumed to be behave as a perfectly competitive market and therefore can be cleared by an optimisation,
- Entry to the capacity market is contestable and free of deterrence, and
- There is no existing capacity

Noting those assumptions, the goal of a formal optimisation model is to find a perfectly competitive equilibrium plant mix where generators build capacity and subsequently sell power in a competitive spot market, which is cleared by an optimisation based on generator offers.

*Optimisation, Social Welfare and Perfect Competition*

Before we present a general optimisation to achieve that goal, we review the relationships between Pareto optimality, optimisation, the maximisation of social welfare, and competitive equilibrium. One measure of economic efficiency is the concept of Pareto-optimality. A feasible allocation $(x^*, y^*)$ is Pareto-optimal if there is no other feasible allocation $(x, y)$ that dominates it, that is one for which one player is better off and no other players worse off. Mathematically the definition is as follows:

Pareto Optimality: The allocation $x*$ is Pareto optimal if there exists no allocation $x$ such that $x_i \succ= x_i^* \forall i$ and $x_j \succ x_j^*$ for some j, where $x* = \left\{x_1^*, x_2^*, ...x_i^*\right\}, x = \left\{x_1, x_2, ...x_i\right\}$ with $x_n^*, x_n \in \mathbb{R}^j$ are allocations of j goods to i consumers.

While Pareto efficiency is a useful concept, there are a number of Pareto-optimal allocations and the concept does not provide any means for selecting the best, or most preferred allocation.

In economic analysis, the traditional approach to addressing this objective has been the maximisation of social welfare and where certain conditions are satisfied, optimisation models have been implemented with this goal in mind. To represent the welfare of society with a single function assumes the adoption of a cardinal utility function, as opposed to a more general ordinal utility function. Here and throughout, we assume that utility can be represented by a cardinal utility function. The maximisation of social welfare is then:

$$\textit{Maximise} \qquad B(x) - C(y) \tag{1.8}$$

$$\textit{Subject to}: \qquad \sum_i x_i^j = \sum_i y_i^j \quad : \lambda^j \qquad \forall j \tag{1.9}$$

$$0 \le y_i^j \le \mathrm{W_i^j}, x_i^j \ge 0 \qquad \forall i \tag{1.10}$$

Where B(x) and C(y) are respectively the social benefit and cost functions and x is as before, an allocation of j goods to i consumers with initial endowments of $\mathrm{W_i^j}$ of good j. We assume the goal of the market mechanism is the maximisation of social welfare. This implicitly assumes the existence of cardinal utility functions. The market clearing constraint ensures that the total initial endowment must match the allocation, while the restriction prevents agents being trading endowments they do not possess. $\lambda^j$ can be considered as the prices of the goods. All players are assumed price takers. Using vector notation, and assuming an interior solution, the first order conditions for a maximum of the social welfare optimisation are:

$$\frac{\partial B}{\partial x} - \lambda = 0 \tag{1.11}$$

$$\frac{\partial C}{\partial y} - \lambda = 0 \tag{1.12}$$

Where an agent/good combination is not an interior solution, it simply means that it was not worth the agent trading the good, or the agents endowment was exhausted while it remains worthwhile to transfer goods away from the agent to another. In equilbrium, the value an additional unit of endowment of a good is equal to the net benefit it brings, Graphically, the marginal change in the social welfare objective can be seen as the difference of two functions:

$$\frac{\partial B}{\partial x} - \frac{\partial C}{\partial y} \tag{1.13}$$

The maximum surplus is obtained when all positive differences between the marginal benefit (demand) function and the marginal cost (supply) function have been exhausted.

In some cases, where demand is either fixed and genuinely independent of price, or where demand responses are included as supply measures, the maximisation of social welfare can be reduced to the minimisation of production cost. This amounts to netting the demand response, if any from the supply-side of the more common "supply=demand" equilibrium definition, as well as avoiding the complications associated with consumer utility functions.

In electricity markets in particular, the cost minimisation concept has been extended significantly through the inclusion of a large number of network constraints that tie many individual markets together to form the basis of the modern system of spot market clearances used in many jurisdictions today. Additionally, the cost minimisation extends to several different types of markets, such as reserve provision, that are inter-related with energy markets, and the technological characteristics of each individual plant are significantly more detailed than in high level studies such as

this. In this study, we use profit maximisation as an organising principle, but under perfect competition, with no control over the market price, the maximisation of profit is equivalent to the minimisation of cost as profit and cost are the sole components of a fixed level of revenue.

In a single commodity case, the maximisation of social welfare corresponds with the more commonly understood "supply = demand" criteria of equilibrium. Under perfect competition, supply and demand can be thought as marginal cost and marginal benefit functions. We can form an optimisation directly by integrating the marginal cost and benefit functions to form a total benefit/cost function, the difference of which defines the total surplus in a market. Multi-commodity equilibriums result when products are linked, for example as substitutes or complements, or as inputs and outputs for one another. Price changes in one commodity move the supply or demand curve of others. In general, it is not the case that we can integrate inverse supply/demand functions to get a consumer surplus function in a multi-commodity case. Nevertheless, in these cases where we cannot define a consistent total surplus function, we can use the KKT conditions to define the equilibrium as it is based on marginal conditions and this is the broad approach taken in many planning models of the input-output nature.

Under perfect competition, and given the ability to represent social welfare with a cardinal function, the notion of Pareto-optimality is perfectly aligned with the concept of optimisation, taking the form of either the maximisation of total surplus, or the minimisation of the total cost of servicing a fixed level of demand. Whenever the marginal benefit of a transaction (to a consumer) exceeds the marginal cost of a transaction (to a supplier), the solution is not Pareto-optimal as there exists a transaction that would be mutually beneficial. Such a transaction would also raise the total surplus, which implies a solution cannot be optimal if it is not Pareto optimal. Alternatively, where the solution is Pareto-optimal, no further beneficial trades exist, implying that the surplus as measured by the difference between the demand and supply curves is maximised.

The equivalence of Pareto-optimality and the maximisation of social welfare implies not only a matching of consumers to producers, where the marginal benefit of the former exceeds the marginal benefit of the latter, but also that the marginal benefit of any consumer trading exceeds the marginal cost of any producer that actually produces. In equilibrium all opportunities for consumers and producers to profitably trade between themselves must be exhausted. Similarly, in a net market, where demand is fixed and the cost of supply is minimised, Pareto-optimality requires that the lowest cost suppliers, including any demand responses, are used to supply the fixed level of demand.

We now consider the relationship between Pareto-efficiency/optimality and competitive, or Walrasian equilibria. A competitive equilibrium is a feasible solution in which every player prefers their position to any other affordable position. Affordability is naturally contingent on an initial wealth distribution, but given a wealth distribution as an initial allocation of commodities, the equilibrium condition for consumers can be stated as: if $\lambda \cdot x_i < \lambda \cdot x_i^*$ then $x_i \leq_i x_i^*$.

This relationship between competitive equilibria and Pareto optimality is summarised by the Fundamental Theorems of Welfare Economics (FTWE). These theorems evolved from ideas promoted by Adam Smith, Pareto, and Barone amongst others, however the first rigorous proof of the theorems is

attributed to Arrow & Debreu (1954). They relate the concept of competitive equilibrium to the Pareto allocation of resources. The Fundamental Theorems respectively state that:

> *Any competitive or Walrasian equilibrium leads to a Pareto-efficient allocation of resources; and,*
>
> *Any Pareto-efficient allocation can be sustained by a competitive equilibrium*

The theorems rely on the assumption of no transaction costs, perfect information, and local non-satiation of preferences. These are minimal assumptions and compatible with the problem definition. The implications of these theorems for analysis of market status, such as in this thesis, is profound. In what follows, it is critical to note that from the first theorem, if an equilibrium is competitive it must feature a Pareto-efficient allocation of resources. Accordingly, if an equilibrium does not exhibit a Pareto-efficient allocation of resources then it must not be a competitive or Walrasian equilibrium. This is important in the context of understanding whether or not various modelling implementations in this area do, or do not, define competitive equilibria. The second theorem is also important. It states that a Pareto-efficient allocation can be sustained by a competitive equilibrium.

The FTWE do not address the case where participants are not price takers. While we still can calculate an equilibrium in these cases and the equilibrium will be Pareto-optimal, an equivalent equilibrium may not be able to be sustained on a competitive as implied by the second theorem. In these cases the relationship between the social optimum and Pareto optimality is broken. In such equilibria there exist socially beneficial trades, where the marginal benefit of the transaction exceeds the marginal cost of the transaction, but these trades do not take place because the gamed equilibrium is Pareto optimal.

In summary, Pareto-optimality and the optimisation of the appropriate surplus/cost measure are analogous under perfect competition. The link between these two enables us to invoke the FTWE in order to consider the status of the equilibriums that are defined by various optimisations. We now turn our attention to the specific situation of investment in generation capacity and the operation of that capacity.

*Two-Stage Formulation*

Before we present a single-stage optimisation we consider the actual decision structure, which contains two stages. Investment and generation are sequential, and not simultaneous, and this is widely acknowledged in the context of research into gaming (Hobbs, Metzler, & Pang, 2000), (Murphy & Smeers, 2005). Under rational expectations, investment decisions are giving consideration to the spot market clearances and PDC implied by those investment decisions. This approach is ultimately not necessary in a perfectly competitive environment, but while the problem need not necessarily be formulated as an MPEC, the problem has the structure required, it can be formulated in this fashion, and it is important to record the precise nature of the problem to clarify all the operations and assumptions that lead to more typical modelling approaches.

When formulated as an MPEC/EPEC problem, the investment decision is parameterised by pricing determined in the spot market:

$$Maximise \quad \sum_i \int_0^1 \left(\lambda^*\left(u;CAP_i\right) - \mathrm{MC_i}\right) GEN_i^*\left(u;CAP_i\right) du - \sum_i \mathrm{FC_i} CAP_i \tag{1.14}$$

$$Subject\,to: \qquad 0 \le GEN_i^*\left(u;CAP_i\right) \le CAP_i \qquad \forall i > 0 \tag{1.15}$$

The optimality conditions of the lower level or second stage market clearance apply at all levels of u and define a market price, $\lambda(u)$ as a function of u:

$$Minimise \qquad \sum_i \mathrm{MC_i} GEN_i\left(u\right) \tag{1.16}$$

$$Subject\,to: \qquad \sum_i GEN_i\left(u\right) = \mathrm{L(u)} \quad : \lambda(u) \tag{1.17}$$

$$GEN_i\left(u\right) \ge 0,\ GEN_i\left(u\right) \in F \qquad \forall i \tag{1.18}$$

Where F is the set of integrable functions on [0,1]. Here u corresponds to the utilisation level expressed as a proportion of the period concerned, $\mathrm{FC_i}$ represents the fixed cost of technology i, and $\mathrm{MC_i}$ is the appropriately scaled constant marginal cost of technology i. The set of technologies includes a notional shortage technology for which capacity is unlimited. $L(u)$ is the LDC as a function of utilisation levels. The model variables are:

- $CAP_i$, the total capacity of technology i, which is continuous; and
- $GEN_i(u)$, the generation of technology i, which is a continuous function of the utilisation level.
- $\lambda(u)$, the spot market price as a function of u and restricted to be an integrable function on [0,1].

The objective function maximises total profit over the period for which the LDC is defined by choosing a capacity, $CAP_i$, and generation function, $GEN_i(u)$, for each technology *i*, and reaping prices $\lambda(u)$.

The fundamental constraint in any electricity investment or market model is the requirement to serve load so that total generation must match load in real time. Recognising the absence of free disposal in the electricity system implies forming this constraint as a strict equality, with the concomitant implication that the associated dual value, which loosely corresponds to the energy price, could be positive or negative in an implementation. Situations with negative prices do arise in real markets for a variety of reasons, such as observed with spring-washer effects around network loops (Hogan et al., 1996), or overnight pricing in markets such as Singapore (E.G. Read, personal communication, 2013), where unit commitment decisions result in negative valued offers. Our investigation does not contain any structure that would provoke the occurrence of negative energy prices, but in the interests of generality we formulate the market clearing constraint as an equality constraint.

*Single Stage Optimisation*

The following single stage optimisation is equivalent to the two-stage optimisation above. It follows from S. Dye (personal communication, 2014). Noting the equivalence of profit maximisation and cost minimisation under perfect competition and that the solution to the lower level problem above is a sub-gradient of the upper level problem, we are able to formulate a cost-minimisation.

The single stage formulation of the problem minimises the total fixed and variable costs required to serve a particular load distribution as represented by an LDC. This formulation is stated in very general terms with no specification of the functional form the load duration curve, and the same requirement for the generation function as before, that it be integrable over [0,1]. Like other straight forward optimisation formulations, this formulation assumes the perspective of a central planner, although we note that the solution to this optimisation problem also characterises the equilibrium of a similarly defined perfectly competitive and contestable market (Arrow & Debreu, 1954):

$$\text{Minimise} \qquad \sum_i \left( \mathrm{MC}_i ENRG_i + \mathrm{FC}_i CAP_i \right) \tag{1.19}$$

$$ENRG_i = \int_0^1 GEN_i(u)\,du \qquad \forall i \tag{1.20}$$

$$\sum_i GEN_i(u) = \mathrm{L}(\mathrm{u}) \tag{1.21}$$

$$GEN_i(u) \le CAP_i \qquad \forall i > 0 \tag{1.22}$$

$$GEN_i(u) \ge 0 \qquad \forall i \tag{1.23}$$

Here $ENRG_i$ corresponds to the total energy produced by technology i. We did not use this intermediate definition in the two-stage problem but include here to improve the exposition of the problem in later sections.

## *1.6 Conventional Optimisation Formulations*

The optimisations in Section 1.5.3 are conceptual. Without restricting the function space that applies to the generation function, resolving the representation of the load function and defining the set of available capacity expansions, there appears to be no way of usefully implementing this formulation. This is equally true of the two stage approach and the single stage approach. Each formulation in Section 1.5.3 must be specialised in order to be implemented, and that means selection of LDC representation, definition of the generation function forms, and definition of the set of feasible capacity expansions.

As we have shown, we are able to address the problem with a single stage optimisation. All implementations we are aware of can be broadly considered implementations of the single stage general optimisation from Section 1.5.3 with the LDC specified and the generation functions restricted in form. For the purpose of our discussion, we denote these broad range of possible implementations of the above optimisations as the "conventional optimisation formulations". Murphy & Smeers (2005)

provides a description of this fundamental approach. Conventional optimisation formulations describe a class of optimisation formulations in which:

- Total investment and operating costs are minimised subject to the requirement to serve load.
- Load is specified in the form of an LDC that is divided in to load classes (or into load slices, although this is significantly less common), where the load class/slice boundaries are fixed.
- Decisions are modelled as single stage.
- They share the attribute that they are specialisations of the general formulation from Section 1.5.3 in which the generation functions are restricted to be, for example, piecewise constant or piecewise linear, with the same breakpoints as the LDC.

The final point is the most significant from the perspective of this thesis. This restriction is implicit and rarely described in the general literature, yet it has potentially serious implications for the solution as we shall show. The enabling assumptions are not without reward. They transform the general formulation of Section 1.5.3 into a problem that can be solved using conventional optimisation techniques such as linear, non-linear and mixed integer programming solution algorithms that are present in conventional solvers. In the rest of this section, we consider some of the more common options when formulating a conventional optimisation. We also show an example of how the general formulation matches a linear programming model when the generation functions are restricted to be piecewise constant with fixed breakpoints.

First, we consider the form of the LDC as well as decide on the most appropriate level of granularity to use from the perspective of accurately representing the LDC. Together, the form which may be piecewise linear, piecewise constant or some other functional form, and the granularity interact to determine the accuracy of the LDC approximation. Those issues are issues of approximation, and while there are more/less desirable approaches, these issues are unavoidable in the sense that all models are approximations of reality.

There are also a number of issues that arise when implementing a conventional optimisation approach that are not associated with approximation, and exist even when we assume the data is a precise representation of reality. In conventional optimisation formulations, the representation of the LDC restricts the functional form of generation functions, and thereby also has implications for the representation of the PDC. Naturally the functional form of the generation functions need not match the form of the LDC so any such restriction is an artificial imposition on the model solution. The implication is that the solutions found by many conventional optimisation formulations are only approximate, even when the data is known and is included in the model with full fidelity.

Approximation is of present in all models, but this form of approximation is different. The nature of the approximation is hidden, and the solutions that result can be internally inconsistent and therefore fall short of describing an equilibrium at all. Inconsistency in an equilibrium model is undesirable, because while an equilibrium model can be predicated on the basis of different views, in this case the inconsistency is based on arbitrary restrictions of generation functions, for which there is no good reason to assume any participant would subscribe, let alone all of them. In this chapter we examine the issue further in the context of perfect competition, but further investigation is required to see how this characteristic influences solutions in more complex situations involving gaming.

### 1.6.1 LDC Representations

In conventional optimisations, LDC representations serve two purposes. Most obviously, they represent the LDC. Less obviously, they define the LDC load classes or slices that are used in the optimisation and implicitly restrict the functional form of the generation curves. In this way, the granularity of the LDC representation is also a determinant of the accuracy of the PDC, thereby affecting investment and the overall cost of satisfying load. We now consider some of the various options available to modellers. Throughout this discussion, we maintain a parallel description of the formulation in terms of basis functions, and the more conventional load class oriented formulation. These approaches are exactly equivalent.

*Piecewise Constant LDC*

The most common form of LDC used is the piecewise constant form, in which the LDC is represented as a step function, with load discretised into classes, each having a constant load requirement. Figure 6 shows how an LDC can be represented in this form.

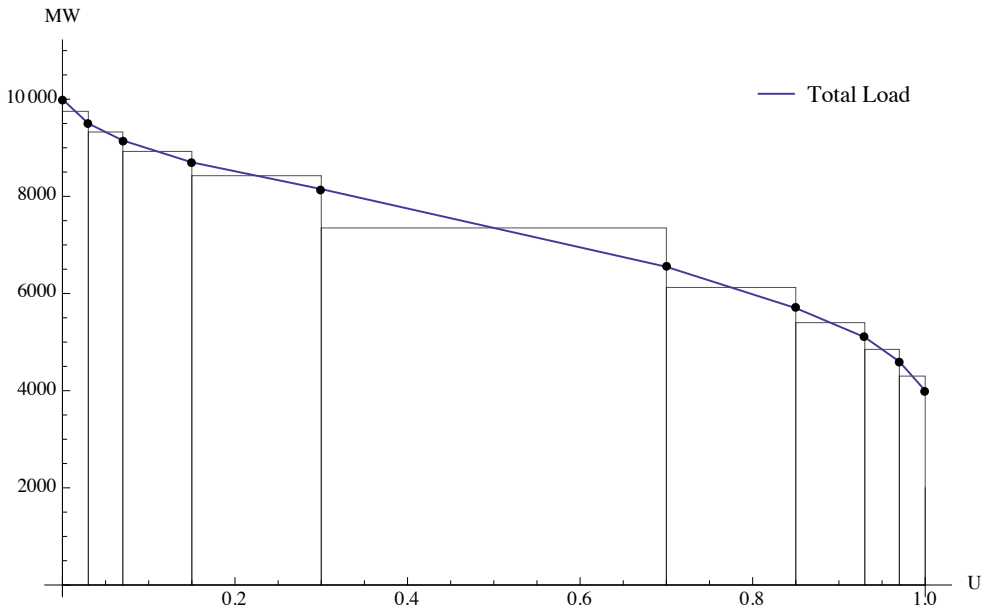



**Figure 6: Piecewise Constant LDC Approximations**

As shown in Figure 6, while a piecewise constant definition of the LDC does represent both the energy and the capacity requirement of the load class, in general we must accept an error in one or other, or both of these measures. The energy requirement of a load class is the area under the relevant section of the LDC. The capacity requirement of a load class is the maximum load across it. In each load class, we require enough capacity to service the actual maximum load level. The energy requirement of the load class is the area under the relevant section of the LDC. If we define load throughout the load class as the actual maximum load level, then the energy content of the load class and hence the variable cost associated with that generation is over-stated, biasing the optimal choice in favour of higher fixed cost/lower variable cost technologies. Furthermore, when energy limits are considered, perhaps in the context of fuel availability or a shortage criteria used by regulators such as in Australia, the mis-specification of the energy content of the load class will lead to either a misallocation of resources or a

solution that is inconsistent with the reality of the constraints. Conversely, if the energy content of a load class is to be represented accurately as shown in Figure 6, the capacity requirements of servicing that load class are understated, resulting in a reversal of the previous bias. As there is no ability to simultaneously match load and energy requirements accurately with a piecewise constant form, this functional form will be inaccurate in one or both dimensions, resulting in non-optimal technological selections as a result of either understating or overstating fixed and variable costs.

Despite this shortcoming, we proceed to specialise the general formulation of the optimal investment problem with a piecewise constant LDC. To do so, we effectively divide the LDC in the fashion shown in Figure 6. To maintain consistency with future formulations we define each load class using utilisation levels $u_k$, where k=0…K, giving a total of K load classes, indexed k=0…K-1. For each load class, load is defined as $L_k$. As described earlier, the representative load level for each load class can be chosen to accurately reflect the energy content, the maximum capacity requirement of the load class, or some intermediate approximation, perhaps designed to minimise a compromise measure of the total error from this form of representation.

We can develop the piecewise constant formulation from the general single stage formulation in Section 1.5.3. We define $P(u_0,\ldots,u_K)$ to be the set of all piecewise linear functions on [0,1] with breakpoints $u_0,\ldots,u_K$. In conventional formulations, these breakpoints are exogenous and those used to define the LDC. We can then replace the very general restriction on generation functions, $GEN_i \in F, \forall i$, with the requirement for those generation functions to be piecewise linear, so that $GEN_i \in P(u_0, \ldots, u_K), \forall i$. Both the generation functions can be written as a weighted sum of a set of basis functions, $b_k$ for k=0,…,K-1. In the case of a piecewise constant representation those are indicator functions which take the value of one in the interval $(u_k,u_{k+1}]$ and zero elsewhere. For later analysis, we also split the energy variable into parts for each load class.

Re-writing the single stage formulation with these adjustments we have:

$$\textit{Minimise} \qquad \sum_i \left( \mathrm{MC}_i ENRG_i + \mathrm{FC}_i CAP_i \right) \tag{1.24}$$

$$\textit{Subject to}: \qquad ENRG_i = \int_0^1 \sum_{k=0}^{K-1} GEN_{i,k} \mathrm{b_k}(\mathrm{u})\, du \qquad \forall i \tag{1.25}$$

$$\sum_i \sum_{k=0}^{K-1} GEN_{i,k} \mathrm{b_k}(\mathrm{u}) = \sum_{k=0}^{K-1} \mathrm{L_k b_k}(\mathrm{u}) \tag{1.26}$$

$$0 \le \sum_{k=0}^{K-1} GEN_{i,k} \mathrm{b_k}(\mathrm{u}) \le CAP_i \qquad \forall i \tag{1.27}$$

$$GEN_i \in P(u_0,\ldots,u_k) \qquad \forall i \tag{1.28}$$

By considering multipliers of each basis function separately and noting $\int_0^1 b_k(u)du = u_{k+1} - u_k$, we can rewrite the general formulation as a linear program in a form that is conventionally used in the literature. As in the general formulation, the objective of the optimal plant mix problem is to minimise

total costs subject to the requirement to serve load and some basic non-negativity conditions relating to generation and capacity variables. The plant mix is "optimal" in the sense that it represents the least total cost method of meeting the load profile, with a given set of technology choices, i=0,…I, ordered in ascending order of fixed cost, where i=0 corresponds to a notional shortage technology but with generation profiles restricted to be piecewise constant with a fixed set of breakpoints. That total cost is represented by the objective function, which requires a class based implementation of the energy cost term to accommodate the partitioning of the LDC into k load classes. The objective function is:

$$\underset{GEN_{i,k},ENRG_{i,k},CAP_i}{Minimise} \quad z = \sum_i \sum_{k<K} \mathrm{MC_i} ENRG_{i,k} + \sum_{i>0} \mathrm{FC_i} CAP_i \tag{1.29}$$

Here $ENRG_{i,k}$ is the total energy generated by technology i in load class k. Here, and throughout, the dual variables corresponding to each constraint are shown on the RHS of each equation, so that mirroring the constraints in (1.20) - (1.23), the equivalent constraints and their accompanying duals are:

$$ENRG_{i,k} - GEN_{i,k}\left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \tag{1.30}$$

$$\sum_i GEN_{i,k} - \mathrm{L_k} = 0 \quad : \lambda_k \qquad \forall k < K \tag{1.31}$$

$$GEN_{i,k} \geq 0 \quad : \varphi^-_{i,k} \qquad \forall i,k < K \tag{1.32}$$

$$CAP_i - GEN_{i,k} \geq 0 \quad : \varphi^+_{i,k} \qquad \forall i > 0, k < K \tag{1.33}$$

$$CAP_i \geq 0 \quad : \chi^-_i \qquad \forall i > 0 \tag{1.34}$$

In each of the load class constraints above, the scope of the constraints is limited to k<K, as k identifies a load class with width $\mathrm{u_{k+1}} - \mathrm{u_k}$. The definition of energy is explicitly stated in (1.30) although in practice this could be substituted into the objective function in order to reduce the dimensionality of the problem. We also note that the dual variable $\lambda_k$, from (1.31), represents the (scaled) market price for load class k and remains theoretically free to take positive and negative values.

The dual constraints corresponding to this problem are:

$$\frac{\partial z}{\partial ENRG_{i,k}} = \mathrm{MC_i} - \varepsilon_{i,k} \geq 0 \qquad \forall i,k < K \tag{1.35}$$

$$\frac{\partial z}{\partial GEN_{i,k}} = -\lambda_k + \varepsilon_{i,k}\left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) + \varphi^+_{i,k} - \varphi^-_{i,k} \geq 0 \qquad \forall i,k < K \tag{1.36}$$

$$\frac{\partial z}{\partial CAP_i} = \mathrm{FC_i} - \sum_k \varphi^+_{i,k} - \chi^-_i \geq 0 \qquad \forall i > 0 \tag{1.37}$$

For those technologies generating in load class k, (1.35) and (1.36) hold with strict equality. We have the following expression relating, $\lambda_k$, the cost to the system of servicing incremental load in load class k to the marginal cost of each technology:

$$\lambda_k = \mathrm{MC_i}\left(\mathrm{u_{k+1}} - \mathrm{u_{k-1}}\right) + \varphi^{+}_{i,k} - \varphi^{-}_{i,k} \qquad \forall i, 0 < k < K \qquad (1.38)$$

When $\varphi^{+}_{i,k} = 0$, implying spare capacity of technology i in load class k, the system cost is the marginal cost of the marginal generator multiplied by the duration of the load class.

$$\lambda_k = \mathrm{MC_i}\left(\mathrm{u_{k+1}} - \mathrm{u_{k-1}}\right) \qquad \forall i, 0 < k < K \qquad (1.39)$$

When $\varphi^{+}_{i,k} > 0$, technology i is infra-marginal so that:

$$\varphi^{+}_{i,k} = \lambda_k - \mathrm{MC_i}\left(\mathrm{u_{k+1}} - \mathrm{u_{k-1}}\right) > 0 \qquad \forall i, 0 < k < K \qquad (1.40)$$

In this case, technology i is inframarginal and earns a profit equal to the difference between the system cost and its own marginal cost of generation for that duration. The final dual constraint collects these inframarginal profits and relates them, with an additional term reflecting the influence of the minimum capacity constraint, to the capital cost recovery required to support investment.

*Piecewise Linear LDC*

As an alternative to the piecewise constant LDC, we may elect to use a different functional form to represent the LDC. That could involve functions of many forms, but here we restrict our analysis to the set of piecewise polynomial functions. The first, and most obvious, extension of this type is the piecewise linear LDC approximation, which permits load and generation to vary linearly, rather than be constant, within a load class. Noting that a piecewise constant LDC is a subset of piecewise linear LDC's, we wish to clarify that in this thesis the term "piecewise linear" should be taken as short-hand for a piecewise linear function that spans the entire range of load levels between minimum and maximum load. Figure 7 shows a linear approximation for a typical load class and contrasts this with the constant approximation with identical energy content.

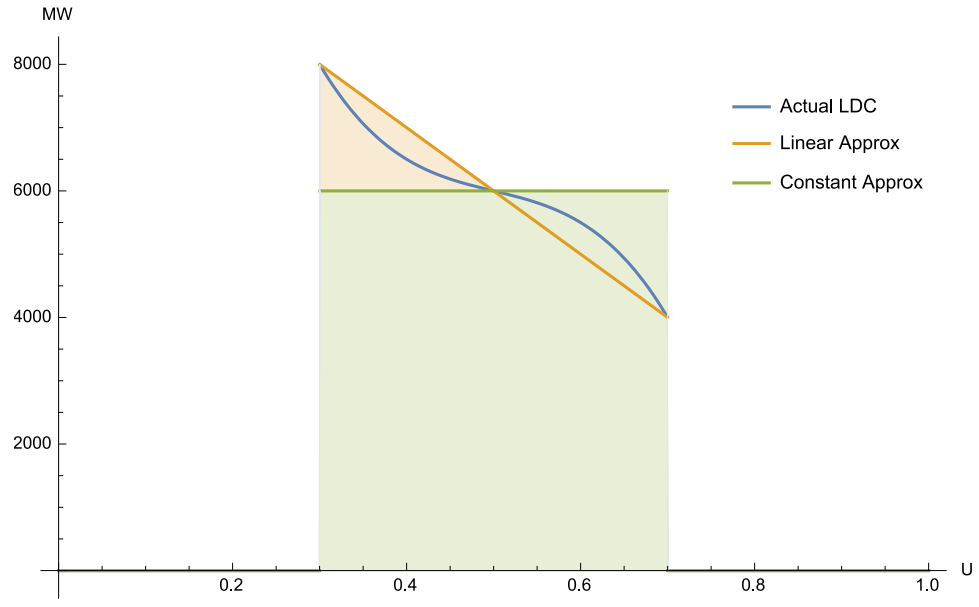



**Figure 7: Comparative LDC Approximations**

We can adapt the general formulation to the piecewise constant case also. The piecewise linear LDC occasions the inclusion of an additional set of basis functions. These take the form of being one from utilisation of zero to its left hand break point, then decreasing linearly to zero at its right hand breakpoint, then continuing as zero up to full utilisation.

$$\textit{Minimise} \qquad \sum_i \left( \mathrm{MC}_i ENRG_i + \mathrm{FC}_i CAP_i \right) \tag{1.41}$$

$$\textit{Subject to}: \qquad ENRG_i = \int_0^1 \sum_{b=0}^{1} \sum_{k=0}^{K-1} GEN_{b,i,k} \mathrm{b}_{\mathrm{b,k}}(\mathrm{u})\, du \qquad \forall i \tag{1.42}$$

$$\sum_i \sum_{k=0}^{K-1} GEN_{b,i,k} \mathrm{b}_{\mathrm{b,k}}(\mathrm{u}) = \sum_{k=0}^{K-1} \mathrm{L}_{\mathrm{b,k}} \mathrm{b}_{\mathrm{b,k}}(\mathrm{u}) \qquad \forall b \tag{1.43}$$

$$0 \le \sum_{b=0}^{1} \sum_{k=0}^{K-1} GEN_{b,i,k} \mathrm{b}_{\mathrm{b,k}}(\mathrm{u}) \le CAP_i \qquad \forall i \tag{1.44}$$

$$GEN_i \in P(u_0, \ldots, u_k) \qquad \forall i \tag{1.45}$$

Corresponding to the new basis functions are a new set of generation variables. To define energy and total capacity use, we must some over the range of basis functions which, in the case of the piecewise linear LDC model, consists of the indices 0 and 1. As before there is an equivalent optimisation to the conventional optimisation formulation in Section 1.5.3. This can be seen in the modification of the load definition. Retaining the load indexing from k=0…K-1, we can specify load with the following function, where $\mathrm{u}_{\mathrm{k}} \le \mathrm{u} \le \mathrm{u}_{\mathrm{k+1}}$ :

$$\mathrm{L}_k(\mathrm{u}) = \mathrm{L}_{0,\mathrm{k}} + \mathrm{L}_{1,\mathrm{k}} \frac{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}}{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}}} \qquad \forall k < K \tag{1.46}$$

Here we have generalised our notation to accommodate higher order forms of LDC representation by specifying by an additional index o=0,1…O, where o reflects the polynomial order of each term in the load specification. In this case $\mathrm{L}_{0,\mathrm{k}}$ represents the constant term or constant basis function, corresponding to load, and $\mathrm{L}_{1,\mathrm{k}}$, the additional peak load attributable to the linear load profile in load class k at utilisation level $\mathrm{u}_{\mathrm{k}}$, which corresponds to a linear basis function.

The selection of $\mathrm{L}_{0,\mathrm{k}}$ and in this case $\mathrm{L}_{1,\mathrm{k}}$ does not necessarily have to coincide with the LDC. The piecewise linear form allows significantly more flexibility in the representation of both energy and capacity requirements within a load class but, like all approximations, will in general still be subject to error. If the piecewise linear form is to respect capacity requirements and be continuous, then we are led to a natural definition of both the constant term and linear coefficient:

$$\mathrm{L}_{0,\mathrm{k}} = \mathrm{L}_{\mathrm{k+1}} \qquad \forall k < K \tag{1.47}$$

$$\mathrm{L}_{1,\mathrm{k}} = \mathrm{L}_{\mathrm{k}} - \mathrm{L}_{\mathrm{k+1}} \qquad \forall k < K \tag{1.48}$$

By virtue of the basis functions available, the generation function applicable to each load profile is piecewise linear. To match the load profile within a load class, we require two constraints to define total generation for each load class, where $GEN_{o,i,k}$ defines the generation of technology I, in load profile o of load class k:

$$\sum_i GEN_{0,i,k} - \mathrm{L}_{0,\mathrm{k}} = 0 \quad : \lambda_{0,k} \qquad \forall k < K \qquad (1.49)$$

$$\sum_i GEN_{1,i,k} - \mathrm{L}_{1,\mathrm{k}} = 0 \quad : \lambda_{1,k} \qquad \forall k < K \qquad (1.50)$$

We do not include more basis functions or profiles than are required to represent the LDC itself. If we did, such profiles would necessarily have to cancel each other out in order to precisely match the LDC profile. It seems unlikely that higher order generation functions would be incentivised in this case and therefore their inclusion provides seemingly little advantage for the cost of additional generalisation, but this is not an area we have investigated. The only other limitation on the generation function in our model is the bounds introduced in the general formulation. Although the minimum generation level may be non-zero or even dynamically dependent on the current system state in reality, in (1.51) we define the minimum bound to be zero for each load profile within each load class. This restriction prevents arbitrage of generation between load profiles within a load class. The maximum bound on generation by technology i within a load class relates to the sum of generation in each load profile contained in that load class. As defined previously and in (1.52), this bound is the installed capacity of technology i.

$$GEN_{o,i,k} \geq 0 \quad : \varphi^{-}_{o,i,k} \qquad \forall o,i,k \qquad (1.51)$$

$$CAP_i - \sum_o GEN_{o,i,k} \geq 0 \quad : \varphi^{+}_{i,k} \qquad \forall i > 0,k \qquad (1.52)$$

Having settled the functional form of the generation function within each load profile and load class, we are able to define the energy use of each technology in each load class by aggregation:

$$ENRG_{i,k} - \left( GEN_{0,i,k} + \frac{GEN_{1,i,k}}{2} \right) \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \qquad (1.53)$$

As we continue to assume each technology has an affine total cost structure, the objective remains as is in (1.29). With the addition of a non-negativity restriction for the capacity of each technology we can state the problem in full:

$$\underset{GEN_{i,k},ENRG_{i,k},CAP_i}{Minimise} \; z = \sum_i \sum_{k<K} \mathrm{MC}_{\mathrm{i}} ENRG_{i,k} + \sum_{i>0} \mathrm{FC}_{\mathrm{i}} CAP_i \qquad (1.54)$$

$$\text{Subject to:} \quad \sum_i GEN_{0,i,k} - \mathrm{L}_{0,\mathrm{k}} = 0 \quad : \lambda_{0,k} \qquad \forall k < K \qquad (1.55)$$

$$\sum_i GEN_{1,i,k} - \mathrm{L}_{1,\mathrm{k}} = 0 \quad : \lambda_{1,k} \qquad \forall k < K \qquad (1.56)$$

$$GEN_{o,i,k} \geq 0 \quad : \varphi^{-}_{o,i,k} \qquad \forall o,i,k < K \qquad (1.57)$$

$$CAP_i - \sum_o GEN_{o,i,k} \geq 0 \quad : \varphi^{+}_{i,k} \qquad \forall i > 0, k < K \qquad (1.58)$$

$$ENRG_{i,k} - \left( GEN_{0,i,k} + \frac{GEN_{1,i,k}}{2} \right) \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \qquad (1.59)$$

$$CAP_i \geq 0 \quad : \chi^{-}_{i} \qquad \forall i > 0 \qquad (1.60)$$

Our approach has the same effect, although we ensure the load function is matched by requiring the aggregate of generation functions for each load profile is satisfied at the peak of each load class.

The first order conditions of this optimisation problem are:

$$\frac{\partial z}{\partial ENRG_{i,k}} = \mathrm{MC_i} - \varepsilon_{i,k} \geq 0 \qquad \forall i,k < K \qquad (1.61)$$

$$\frac{\partial z}{\partial GEN_{o,i,k}} = -\lambda_{o,k} + \varepsilon_{o,k} \left( \frac{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}}}{o+1} \right) + \varphi^{+}_{i,k} - \varphi^{-}_{o,i,k} \geq 0 \qquad \forall o,i,k < K \qquad (1.62)$$

$$\frac{\partial z}{\partial CAP_i} = \mathrm{FC_i} - \sum_k \varphi^{+}_{i,k} - \chi^{-}_{i} \geq 0 \qquad \forall i > 0 \qquad (1.63)$$

The application of the dual constraint (1.61) is as before. For technologies generating in load class k, (1.61) holds with strict equality however, unlike the piecewise constant case, there is no longer a unique load profile so (1.62) will hold with equality only for those technologies generating to satisfy load profile o, in load class k. For a technology servicing the baseload profile in load class k we have the following expression relating, $\lambda_{0,k}$, the cost to the system of servicing incremental baseload in load class k to the marginal cost of each technology:

$$\lambda_{0,k} = \mathrm{MC_i} \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) + \varphi^{+}_{i,k} - \varphi^{-}_{0,i,k} \qquad \forall i,k < K \qquad (1.64)$$

Similarly, for technologies servicing the peaking, or linear, profile, we have the following expression:

$$\lambda_{1,k} = \mathrm{MC_i} \left( \frac{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}}}{2} \right) + \varphi^{+}_{i,k} - \varphi^{-}_{1,i,k} \qquad \forall i,k < K \qquad (1.65)$$

In each case the marginal cost is scaled by the average utilisation of capacity across for the appropriate load profile. Therefore, when $\varphi^{+}_{i,k} = 0$, implying spare capacity of technology i in load class k, the system cost is the marginal cost of the marginal generator scaled by the average utilisation of technology i which is defined by the width of the load class and the degree of the polynomial term associated with load profile o in load class k.

$$\lambda_{o,k} = \mathrm{MC_i} \left( \frac{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}}}{o+1} \right) \qquad \forall o,k < K \qquad (1.66)$$

Both $\lambda_{0,k}$ and $\lambda_{1,k}$ correspond to prices for increments in load but relate to different profiles. The first dual variable represents the dual associated with an equal increase in load across the entire load class, whereas the second is associated with an increase in load with a linear profile as shown below in Figure 8:

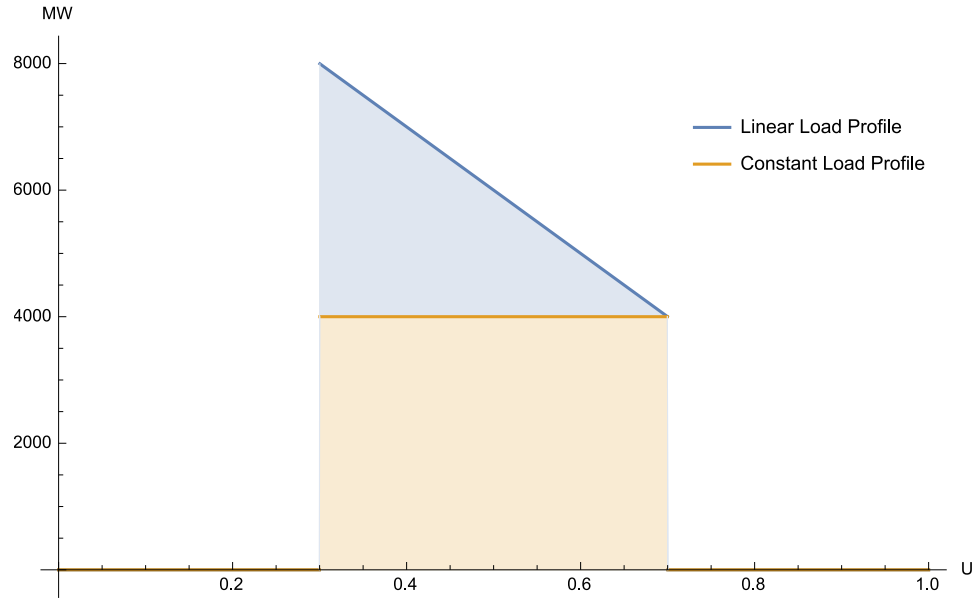



**Figure 8: Piecewise Linear Load Profiles & Pricing**

From the perspective of basis functions, the above profiles represent the constant and linear basis functions in the relevant utilisation range. It follows that the average utilisation of capacity servicing the baseload profile is higher than the average utilisation of capacity servicing the peaking, or linear, profile. As a result, this approach can yield two distinct marginal technologies, one for each profile or basis function, in the same load class. As shown in the dual equations, the associated pricing is, within a scaling, factor determined for each. This may suggest some ambiguity in terms of market pricing, but in this particular instance we are addressing the piecewise linear load formulation and there is only one technology that is marginal across the load class with respect to market pricing, that being the technology servicing the linear load profile. Accordingly, we can determine the market price across the load class as:

$$MP_k = \frac{2\lambda_{1,k}}{\mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k\text{-}1}}} = MC_i \qquad \forall k < K \qquad (1.67)$$

Although the market price can be made clear ex post, the prices are specified in terms of the load class duration and therefore still require scaling. While this can be achieved in an ex-post fashion with relative ease, the necessity of scaling prices is problematic when modelling extensions such as demand response which must be dealt with endogenously to ensure the correct representation of price effects is achieved. The need for scaling is a result of the underlying structure of the formulation, which does not deal with individual periods as the market does, and therefore does not easily produce prices for commodities that correspond with those normally traded in energy markets.

We consider the dual equation (1.63) associated with the investment, or capacity, decision. When $\varphi^{+}_{i,k} = 0$, technology i is infra-marginal so that:

$$\varphi^{+}_{i,k} = \lambda_{o,k} - \mathrm{MC_i}\left(\frac{\mathrm{u_{k+1}} - \mathrm{u_{k-1}}}{o+1}\right) > 0 \qquad \forall o,i,k < K \qquad (1.68)$$

In this case, technology i earns a profit equal to the difference between the system cost and its own marginal cost of generation for that duration. The final dual constraint collects these infra-marginal profits and relates them, with an additional term reflecting the influence of the minimum capacity constraint, to the capital cost recovery required to support investment.

Finally, we note the relationship in (1.68) is valid for all load profiles within load class k, confirming the following relationship between system prices:

$$\lambda_{1,k} = \lambda_{0,k} - \mathrm{MC_i}\left(\frac{\mathrm{u_{k+1}} - \mathrm{u_{k-1}}}{2}\right) \qquad \forall o,i,k < K \qquad (1.69)$$

As we would expect, the cost to the system of an increment in the linear profile is less than and, to be precise, half the cost to the system of an increase in the baseload profile.

*Higher Order Load Class Formulations*

Where non-linearity exists in individual load classes it may be desirable to gain a better compromise between respecting load and energy requirements. In such cases it may be preferable to define a higher order approximation rather than persist with a piecewise linear approach. A higher order functional form, involves the use of additional basis functions in the LDC representation. The advantage of higher order functional forms in this context is the ability to reduce the number of segments required to attain a given level of accuracy in the LDC representation (Dye, 1994). Alternatively, we can state that higher order approximations, such as the piecewise linear approximation, are significantly more accurate than lower order approximations, such as the piecewise constant approximation, given identical levels of granularity. In the limit, given a polynomial of sufficient order, the entire LDC could be well represented in a single load class. However, despite the advantage of higher order representations in this respect, this does not necessarily translate into an overall computational advantage, as handling higher order functions may be more computationally intensive.

Furthermore, we also require accurate representation of the PDC, so that investment incentives are correctly assessed and lead to appropriate investment decisions. If this is not the case, certain aspects of the system's behaviour will be misrepresented or, in the case of a coarse discretisation, may not be represented at all. So while a relatively small subset of utilisation points may result in a piecewise linearisation of acceptable accuracy when measured against the actual LDC, that subset may not provide a satisfactory solution quality. This suggests that the benefits of higher order representations may not accrue if a significant number of utilisation levels are required for the purposes of the overall model anyway. Ideally, we would like to decompose these twin requirements so that the LDC representation could be considered separately from the demands of the model.

For a given load class the functional form of the LDC can be generalised to:

$$L_k(u) = \sum_{o=0}^{O} L_{o,k} \left( \frac{u_{k+1} - u}{u_{k+1} - u_k} \right)^o \qquad \forall k < K \qquad (1.70)$$

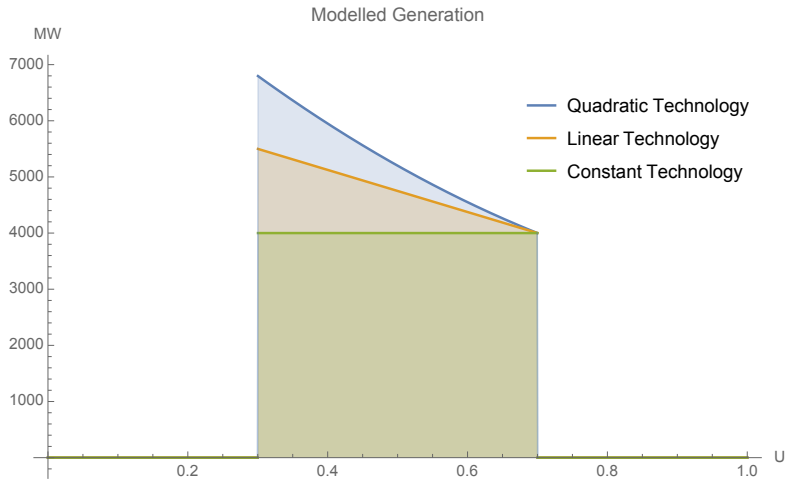
Similarly, for each active term in the polynomial representation we have a generation constraint requiring that load to be met:

$$\sum_i GEN_{o,i,k} - L_{o,k} = 0 \quad : \lambda_{l,k} \qquad \forall o,k < K \qquad (1.71)$$

We also have a more general expression for the energy output of technology i, in load profile o, within load class k:

$$ENRG_{i,k} - \sum_{o=1}^{O} \left( \frac{GEN_{o,i,k}}{o+1} \right) \left( u_{k+1} - u_k \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \qquad (1.72)$$

As before, for each basis function that is introduced we must also consider an additional marginal technology corresponding to that basis function. That is, each basis function could correspond to a separate marginal technology. The example shown in the LHS pane of Figure 9 corresponds to the implementation of a quadratic representation in which three distinct marginal technologies have been selected to serve the three load profiles in the load class. The nature of the decomposition dictates that these are modelled as operating simultaneously, as was the case in the piecewise linear implementation. This is inconsistent with dispatch and market clearing principles which suggest that the load would be served as shown in the RHS panel of Figure 9 given the capacity selection made.



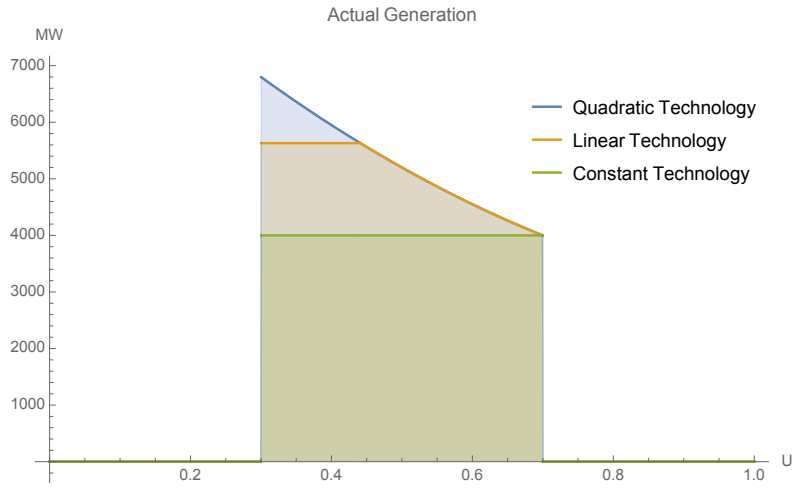



**Figure 9: Generation with Higher Order Load Formulations, Modelled vs Actual generation**

*Alternative Formulation*

As we have seen, the load class approach effectively creates a set of load profiles within each load class, each with different characteristics relating to the particular basis function to which they correspond. The piecewise constant approach involves a single load profile and yields a unique marginal technology and market price. Unfortunately, this approach is not desirable as it cannot simultaneously account for energy and capacity requirements within a load class. The piecewise linear approach resolves two load profiles, and produces two marginal technologies in general but, as one is the constant profile with fixed capacity requirements, we are able to resolve a unique market price and the solution is consistent with market clearing across the load class. Higher order forms offer improved fitting capability but result in market marginal technologies that have spare capacity. These outcomes that are inconsistent with basic merit order market clearing principles. We now introduce an alternative formulation that defines a single marginal technology for each load class, and thereby frees the modeller to use higher order LDC representations if they desire.

One alternative approach involves the consideration of K+1 utilisation and load levels, indexed by k=0,…K, in ascending order of utilisation, which we treat as separate market clearing instances. In the case of adjacent points, these utilisation levels also define the boundary of a traditional load class. As in previous formulations, the objective of the optimal plant mix problem is to minimise total costs subject to the requirement to serve load and some basic non-negativity conditions. The objective, discretised by load class, remains:

$$\underset{GEN_{i,k},ENRG_{i,k},CAP_i}{Minimise} \quad \sum_i \sum_{k<K} \mathrm{MC}_i ENRG_{i,k} + \sum_{i>0} \mathrm{FC}_i CAP_i \tag{1.73}$$

The fundamental constraint is the requirement to serve load so that total generation must match load in real time. We define the constraint directly at a point, defined by $\mathrm{u}_\mathrm{k}$ .

$$\sum_i GEN_{i,k} - \mathrm{L}_\mathrm{k} = 0 \quad : \lambda_k \qquad \forall k \tag{1.74}$$

Generation limits are as before:

$$GEN_{i,k} \geq 0 \quad : \varphi^{-}_{i,k} \qquad \forall i,k \qquad (1.75)$$

$$CAP_i - GEN_{i,k} \geq 0 \quad : \varphi^{+}_{i,k} \qquad \forall i > 0,k \qquad (1.76)$$

The energy generated by each technology i for each load class k is the area of the trapezium formed by the appropriate generation function for the load class. It is set by the constraint below:

$$ENRG_{i,k} - \left( \frac{GEN_{i,k+1} + GEN_{i,k}}{2} \right) \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \qquad (1.77)$$

With the addition of a non-negativity restriction for the capacity of each technology we can state the problem in full:

$$\underset{GEN_{i,k}, ENRG_{i,k}, CAP_i}{Minimise} \qquad \sum_i \sum_{k<K} \mathrm{MC_i} ENRG_{i,k} + \sum_{i>0} \mathrm{FC_i} CAP_i \qquad (1.78)$$

$$\text{Subject to:} \quad \sum_i GEN_{i,k} - \mathrm{L_k} \geq 0 \quad : \lambda_k \qquad \forall k \qquad (1.79)$$

$$GEN_{i,k} \geq 0 \quad : \varphi^{-}_{i,k} \qquad \forall i,k \qquad (1.80)$$

$$CAP_i - GEN_{i,k} \geq 0 \quad : \varphi^{+}_{i,k} \qquad \forall i > 0,k \qquad (1.81)$$

$$ENRG_{i,k} - \left( \frac{GEN_{i,k+1} + GEN_{i,k}}{2} \right) \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i,k < K \qquad (1.82)$$

$$CAP_i \geq 0 \quad : \chi^{-}_{i} \qquad \forall i > 0 \qquad (1.83)$$

The first order conditions of this problem are:

$$\frac{\partial z}{\partial ENRG_{i,k}} = \mathrm{MC_i} - \varepsilon_{i,k} \geq 0 \qquad \forall i,k < K \qquad (1.84)$$

$$\frac{\partial z}{\partial GEN_{i,k}} = -\lambda_k + \frac{\varepsilon_{i,k}}{2} \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) \Big|_{k<K} + \frac{\varepsilon_{i,k-1}}{2} \left( \mathrm{u}_{\mathrm{k}} - \mathrm{u}_{\mathrm{k\text{-}1}} \right) \Big|_{k>0} + \varphi^{+}_{i,k} - \varphi^{-}_{i,k} \geq 0 \qquad \forall i,k \qquad (1.85)$$

$$\frac{\partial z}{\partial CAP_i} = \mathrm{FC_i} - \sum_k \varphi^{+}_{i,k} - \chi^{-}_{i} \geq 0 \qquad \forall i > 0 \qquad (1.86)$$

When technology i is generating at utilisation level $\mathrm{u_k}$, (1.84) and (1.85) hold with strict equality so that:

$$-\lambda_k + \frac{\mathrm{MC_i}}{2} \left( \mathrm{u}_{\mathrm{k+1}} - \mathrm{u}_{\mathrm{k}} \right) \Big|_{k<K} + \frac{\mathrm{MC_i}}{2} \left( \mathrm{u}_{\mathrm{k}} - \mathrm{u}_{\mathrm{k\text{-}1}} \right) \Big|_{k>0} + \varphi^{+}_{i,k} - \varphi^{-}_{i,k} = 0 \qquad \forall i,k \qquad (1.87)$$

For $k \in C, k < K$ and with $\varphi^{+}_{i,k} = 0$ implying spare capacity, we have the following expression relating the system energy price to the marginal cost of the marginal generator:

$$\lambda_k = \frac{\mathrm{MC_i}}{2}\left(\mathrm{u_{k+1}} - \mathrm{u_{k-1}}\right) \qquad \forall i, 0 < k < K \qquad (1.88)$$

Importantly, the marginal price is unique, and barring coincidental equality between cost structures at the relevant utilisation level, the marginal technology is also. The nature of the commodity to which this price relates is somewhat unintuitive. Incremental load at $\mathrm{u_k}$ requires additional generation by the marginal technology at that point. In capacity terms this is costless, as we have spare capacity. However, an increase in load creates additional energy requirements in two load classes, which by virtue of the assumed linear adjustment of the generation function implies the cost is of the form above. This is expressed graphically in Figure 10. Post solution, we can scale prices by the inverse of the utilisation range, to arrive at a market-clearing price for energy. In the case where spare capacity is available this scaling factor is $2/\left(\mathrm{u_{k+1}} - \mathrm{u_{k-1}}\right)$.

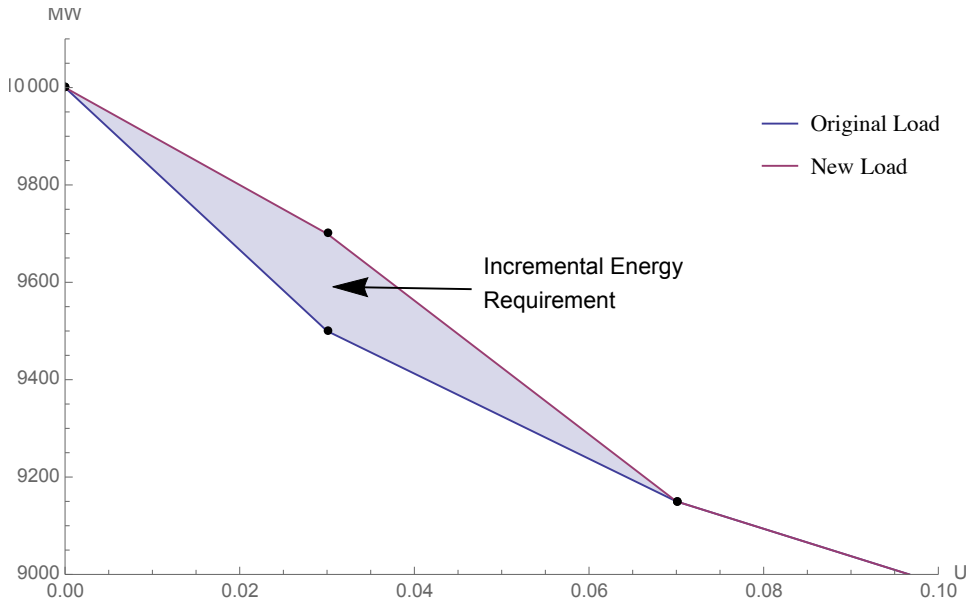



**Figure 10: Pricing in Piecewise Linear Model**

While this formulation is successful in eliminating the issue of multiple marginal technologies and the market clearing inconsistencies that accompany these, the downside of this approach is that it forces the requirement of a single marginal technology when, in reality, the cost of increasing generation according to the pattern shown might involve multiple technologies. This may confuse adjustment of other generation as capacity increases would affect multiple load classes. Issues such as demand response only further complicate consideration as the prices that drive demand response must be dealt with endogenously, rather than post-solution, to ensure the correct representation of price effects is achieved.

### 1.6.2 Conventional Optimisation with Piecewise Linear LDC

The purpose of the following example is to illustrates the properties and inconsistency of conventional optimisation formulations. In doing so, we also illustrate the structural properties of optimal solutions

to the general model and, by comparison, how these two approaches differ. The example is based on the piecewise linear LDC implementation of the conventional approach. We define load classes or basis functions that correspond to the piecewise linear segments of the LDC, so that load varies linearly over each load class. To avoid contaminating the comparison with approximation errors we assume that the LDC is actually piecewise linear in reality and precisely represented. Accordingly, load is linearly interpolated between the points in Table 1. The LDC can be seen later in Figure 11.

| Load (MW) | 90000 | 82000 | 78000 | 58000 | 50000 |
|---|---|---|---|---|---|
| Utilisation | 0.000 | 0.050 | 0.100 | 0.900 | 1.000 |

**Table 1: Hypothetical LDC Definition**

In addition, we have the following technological options with the accompanying cost structures. The costs are not necessarily realistic, although this matters little as the efficacy of a formulation should not be data dependent. Fixed costs are presented in terms of a per unit charge, while variable costs are scaled to reflect the variable cost of operating continuously for the entire time period, in this case a year:

| Technology | Fixed Cost | Variable Cost |
|---|---|---|
| Notional Shortage | 0 | 15000 |
| OCGT | 50 | 3500 |
| CCGT | 220 | 2000 |
| Coal | 650 | 500 |
| Geothermal | 1000 | 0 |

**Table 2: Hypothetical Technological Cost Structure**

The following model, as presented in Section 1.6.1, is applicable to this data structure:.

$$\underset{GEN_{i,k},ENRG_{i,k},CAP_i}{Minimise} \quad \sum_i \sum_{k<K} \mathrm{MC_i} ENRG_{i,k} + \sum_{i>0} \mathrm{FC_i} CAP_i \tag{1.89}$$

$$\text{Subject to:} \quad \sum_i GEN_{i,k} - \mathrm{L_k} \geq 0 \quad : \lambda_k \qquad \forall k \tag{1.90}$$

$$GEN_{i,k} \geq 0 \quad : \varphi^-_{i,k} \qquad \forall i,k \tag{1.91}$$

$$CAP_i - GEN_{i,k} \geq 0 \quad : \varphi^+_{i,k} \qquad \forall i > 0, k \tag{1.92}$$

$$ENRG_{i,k} - \left( \frac{GEN_{i,k+1} + GEN_{i,k}}{2} \right) \left( \mathrm{u_{k+1}} - \mathrm{u_k} \right) = 0 \quad : \varepsilon_{i,k} \qquad \forall i, k < K \tag{1.93}$$

$$CAP_i \geq 0 \quad : \chi^-_i \qquad \forall i > 0 \tag{1.94}$$

The variables are the capacity of each technology, $CAP_i$ and generation by each technology at each utilisation level k, $GEN_{i,k}$. The duration of each load class k is $u_{k+1} - u_k$ so that for K utilisation levels there will be K-1 load classes.

The optimal capacity mix prescribed by this optimisation is:

| Technology | Capacity |
|---|---|
| Notional Shortage | 0 |
| OCGT | 12000 |
| CCGT | 0 |
| Coal | 20000 |
| Geothermal | 58000 |

**Table 3: Optimisation Solution of Conventional Optimisation Formulation**

The total cost of building capacity and servicing load with this plant mix is 78.35M.

*Optimality*

By assumption, we have abstracted away from the specific issues surrounding the degree of approximation that is introduced by a functional LDC representation. It is tempting therefore to think the formulation, having a precise representation of reality, will produce an optimal solution. Unfortunately, this is not the case. Naturally the solution is optimal for the formulation specified by (1.68)-(1.73), however it is not optimal for the general model, as specified in Section 1.5.3, which represents the actual problem in reality. The actual optimal solution is :

| Technology | Conventional Approach | Thesis / Screening Curve Solution |
|---|---|---|
| Notional Shortage | 0 | 696 |
| OCGT | 12000 | 11637 |
| CCGT | 0 | 4333 |
| Coal | 20000 | 10333 |
| Geothermal | 58000 | 63000 |
| Optimal Solution Total Cost | 78.35M | 77.43M |

The total cost of capacity and energy generation is 77.43M. This solution was arrived at using the approach detailed in Chapter 2, but in this case the problem collapses to a simpler problem. As the set of generation function breakpoints includes the utilisation levels corresponding to optimal trade-offs, the solution is Pareto-optimal. Given generators are price takers by assumption, a Pareto optimal solution must also represent cost minimisation of the total cost function. No generation could profitably, or at lower cost, be switched from one technology to another. In any case, although the solution using the screening curve/thesis approach is optimal, for the purpose of this comparison, and

to illustrate the solution obtained using the conventional optimisation formulation is non-optimal, it suffices to show that the solution was bettered. The total cost of supply is lower than that obtained by the conventional optimisation of 78.35M, proving the conventional approach is sub-optimal and not a solution to the general formulation presented in Section 1.5.3.

*Equilibrium status*

We now show the conventional optimisation formulation does not describe a competitive equilibrium. The first Fundamental Theorem states that any competitive or Walrasian equilibrium leads to a Pareto-efficient allocation of resources. If the conventional formulation describes a competitive equilibrium then the FTWE dictates that the solution will also represent a Pareto-efficient allocation of resources. To show the solution is not a competitive equilibrium, it suffices to find a single example of a Pareto-efficient trade that would be mutually beneficial.

In the solution to the conventional optimisation formulation, the CCGT investor does not build any capacity, and therefore has a cost of $0. The coal generator constructs capacity of 20,000, and operates at full capacity with utilisation of 0.1, with generation dropping linearly to zero from that utilisation level to a utilisation level of 0.9, beyond which point load occurring with greater frequency is serviced by geothermal generation. The total cost of this operation is 20,000x650 + 10,000x500=$18M.

Suppose the coal investor pays the CCGT $0.5M to install 1000 units of CCGT capacity, so that the coal investor can reduce capacity by 1000. The new solution sees coal capacity of 19,000 and CCGT capacity of 1,000. The CCGT investor services the incremental load from L(0.14) = 77,000 to L(0.1)=78,000. This represents the last 1,000 units of capacity required by the coal investor. The CCGT investor performs this operation at a cost of 1,000x200+120*2000=$0.24022M. Offsetting this, they receive $0.5M from the coal investor so they are better off by $0.25978M. The coal investor now has total cost of 19,000x650+9880x500=$17.29M, which represents a saving of $0.71M. After compensating the CCGT investor, the coal investor is better off by $0.21M.

Since the coal investor and the CCGT are better off, the conventional optimisation formulation is not a Pareto optimal solution. Therefore, by the FTWE, it cannot describe a competitive equilibrium of the problem under examination. To be clear, the solution is an equilibrium to a restricted system in which all generation is restricted to follow a piecewise linear generation duration curve with the same breakpoints as the LDC but that is not the system under study.

*LDC Representation and Generation Restrictions*

From the narrow perspective of seeking to accurately represent the LDC, we have identified the relative benefits of various functional forms in their own right. In particular, we have pointed out the need for non-constant representations if the twin requirements of representing capacity and energy content within a load class are to be satisfied. And we noted that, while higher order functional forms enable more faithful representation of the LDC, quadratic and higher forms can, if the generation function structures are overly restricted, result in outcomes that are inconsistent with market clearing principles. The formulation above uses a piecewise linear LDC form and as a result of the actual LDC being piecewise linear, the conventional optimisation above represents the capacity and energy requirements of each load class precisely. In this respect, the conventional optimisation formulation

represents the actual LDC perfectly, by assumption. But even if that assumption were removed there would be no advantage gained by any method presented in this thesis, as those methods would use the same representation of the LDC.

However, as we have shown, when implementing a conventional optimisation the LDC definition has implications for the structure of the optimal generation functions. The breakpoints that form part of the definition of the optimal generation function are restricted in conventional optimisation formulations to those break points that define the LDC. The generation profiles that are defined by the solution to the conventional optimisation formulation are evident in the LDC filling in Figure 11.

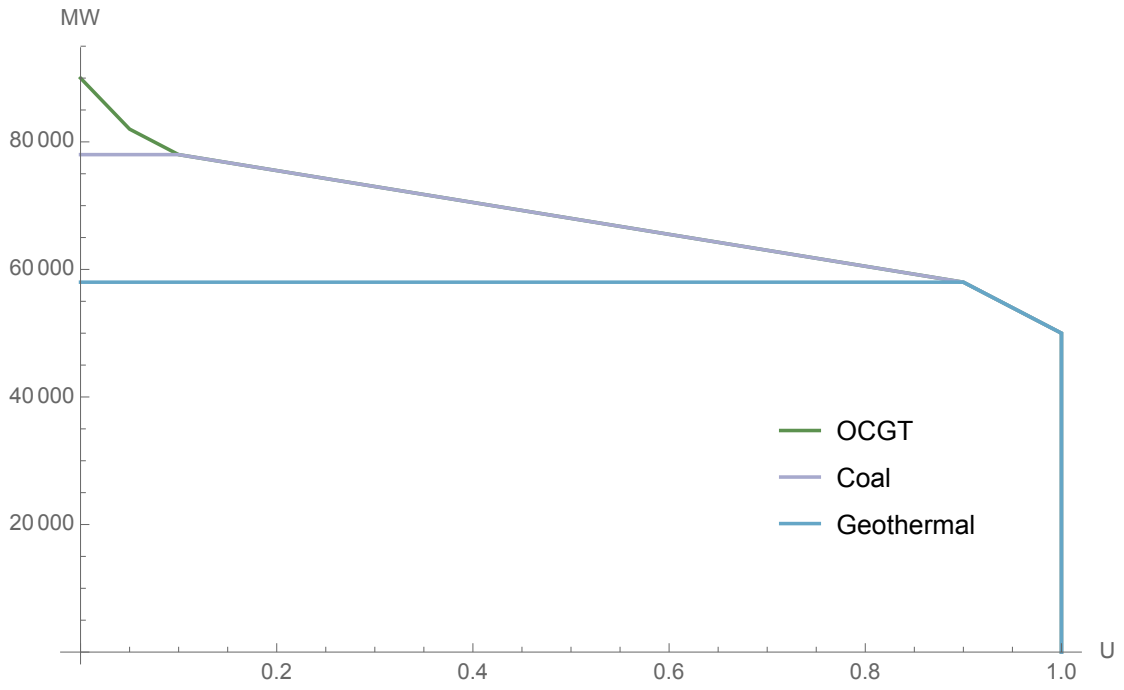



**Figure 11: LDC Filling for Conventional Optimisation Solution**

Note that the breakpoints used in the generation profiles are the same as those that define the LDC. By comparison, the LDC filling of the optimal solution is in Figure 12. In Figure 12, it is notable that unlike the solution to conventional optimisation problem in Figure 11, the generation functions for each technology do not have breakpoints that coincide with the breakpoints necessary for definition of the LDC.

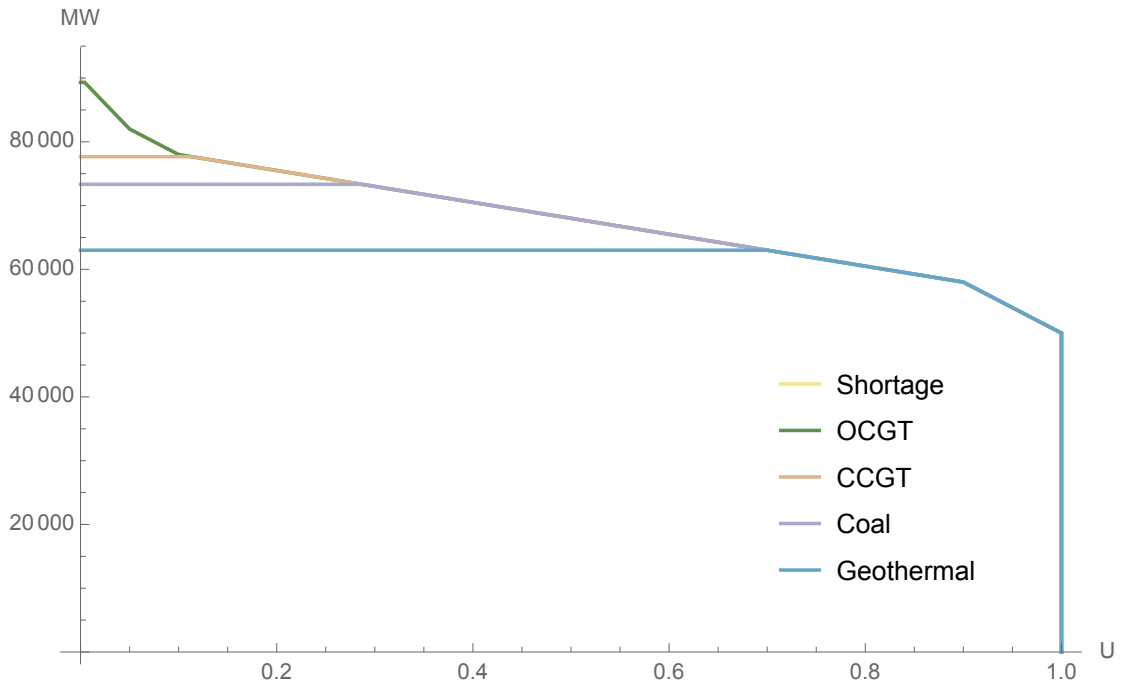



**Figure 12: LDC Filling for Optimal Solution**

For example, in the solution to conventional optimisation shown in Figure 11, the generation function of geothermal peaks at a utilisation level of 0.9, whereas in the generation function corresponding to the actual optimal solution as shown in Figure 12, geothermal generation peaks at a utilisation level of 0.7, which is not a level at which the LDC is defined.

*PDC Representation*

As discussed in Section 1.6.1, the recovery of pricing from models involving more than a single basis function for representing load within each load class is more complex than in the piecewise-constant case, which involves only a single, constant, basis function for load in each load class. The optimal PDC is based on market clearances arising from the optimal capacity from the actual solution. In this simple case, the PDC corresponds also to the screening curve solution, in which the marginal cost of the marginal technology is price setting. The PDC corresponding to the optimal solution is shown below in Figure 13 (note the price axis is truncated at 5000, whereas shortage costs actually rise to 15,000).

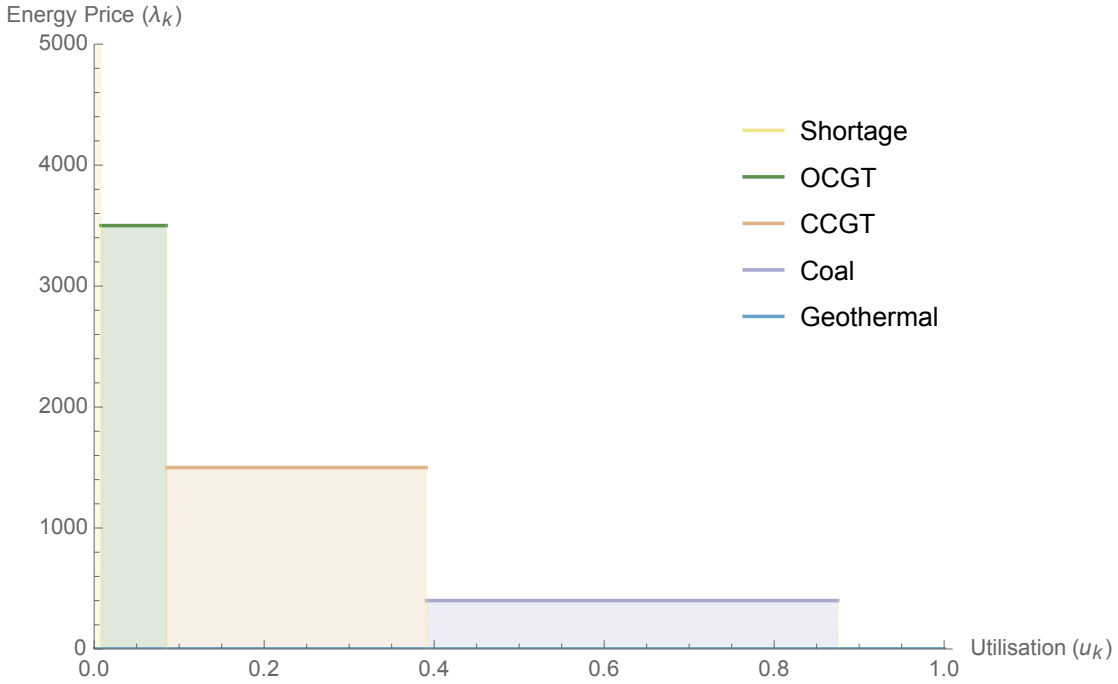



**Figure 13: Optimal PDC**

This PDC provides cost recovery precisely for each technology, and the prices correspond to marginal costs that occur naturally as part of the spot market clearing process. Consider the optimal cost recovery of CCGT for example. CCGT earns a profit of 13,000 during times of shortage which occurs with a relative frequency of 0.00435 (5dp). In addition CCGT earns a profit of 1,500 when OCGT is marginal, which occurs with a relative frequency of 0.10899 (5dp). In total then, cost recovery of $56.52 is available from shortage periods, and $163.48 when OCGT is marginal, which equates to a total of $220, the fixed cost of CCGT per unit of capacity.

*Solution Consistency*

In this example, the difference between the optimal objective function value of the conventional approach and the precise solution is relatively small. However the solutions are significantly different when viewed from the perspective of consistency.

To examine the consistency of the solution to the conventional optimisation, we consider the PDC that corresponds to the market clearances that would result from application of the market clearing model using the capacity values determined by the conventional optimisation approach. This PDC is shown below:

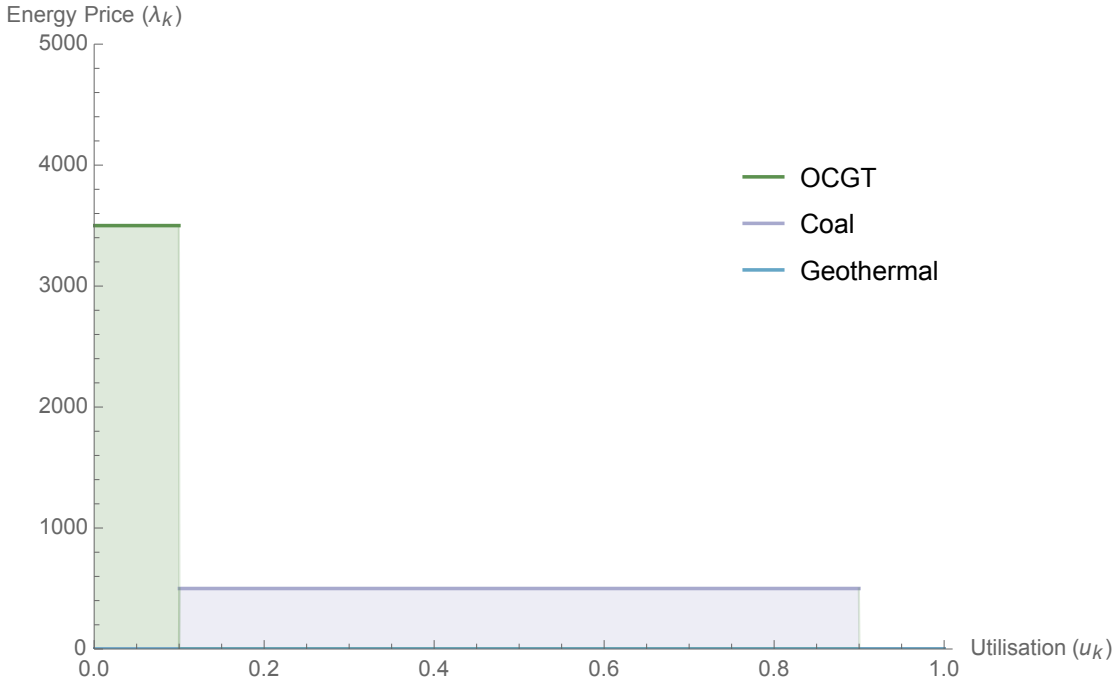



**Figure 14: Spot Market Consistent PDC**

This is the PDC that investors with rational expectations believe will result from perfectly competitive spot market clearances given the capacity investment prescribed by the conventional approach. This PDC does not provide cost recovery so we have the situation where the capacity level is incompatible with the spot market outcome. For example, this PDC implies that OCGT will be marginal 10% of the time, and there will be no shortage. As a result there is no opportunity for OCGT investors to recover installation costs as while OCGT is the marginal technology, the spot market price is $3,500 which is equal to the marginal cost for this technology.

The spot market clearing process specified in Section 1.5.3 does not provide for the inclusion of uplift pricing components, and the PDC which the conventional approach relies upon to support investment is not compatible with rational expectations when investors know the structure of the spot market.

In the conventional optimisation formulations, spot market pricing discipline comes explicitly, as opposed to implicitly, from the investment constraint. This is not the case in the problem described by the general formulation, in which the cost of capacity installation only influences spot market pricing through its influence on market structure. Although some markets have rules designed to support cost recovery of peaking plant, and in general markets are designed to implicitly result in recovery of costs, we have yet to find a market clearing procedure in any electricity market that accounts for cost recovery when setting the spot market price at all levels of load, in the general course of market clearance.

Accordingly, rational investors will not invest in OCGT technologies given an expectation of the PDC shown in Figure 14 because OCGT is never infra-marginal and therefore cannot ever contribute to its own cost recovery. The same is true for every other technology in the "optimal" mix prescribed by the conventional optimisation formulation, as investors in lower marginal cost

technologies will be aware that their infra-marginal profit opportunities will not necessarily materialise as no technology will be built to the capacity predicted by the model.

As we have stated the reason for this inconsistency is the artificial restriction on generation functions in conventional optimisation formulations. Therefore, in a technical sense, for an investor to invest on the basis of the cost recovery PDC from the conventional optimisation formulation, they must assume that all other investors, present and future, will perceive the same generation function restrictions that they do. If other investors do not, and they should not if they are rational, their investment patterns will differ and cost recovery will not be achieved.

The requirement for uplift pricing components in conventional optimisation formulations is a general phenomenon among conventional formulations rather than specific to the piecewise linear formulation. For example, if we examine the dual constraints corresponding to the piecewise-constant LDC variant of the optimisation problem:

$$\frac{\partial z}{\partial GEN_{i,k}} = -\lambda_k + \mathrm{MC_i} + \varphi^+_{i,k} - \varphi^-_{i,k} \geq 0 \qquad \forall i,k < K \tag{1.95}$$

$$\frac{\partial z}{\partial CAP_i} = \mathrm{FC_i} - \sum_k \left(\mathrm{u_{k+1}} - \mathrm{u_k}\right)\varphi^+_{i,k} - \chi^-_i \geq 0 \qquad \forall i > 0 \tag{1.96}$$

When $GEN_{i,k} > 0$, we can express $\varphi^+_{i,k}$ in terms of system prices and marginal costs at utilisation levels $\mathrm{u_k} < u_{k*}$, where $u_{k*}$ is the utilisation level at which at technology i enters the optimal plant mix, so that $\varphi^+_{i,k} = \lambda_k - \mathrm{MC_i}$, so that:

$$\mathrm{FC_i} - \sum_{k \leq k*} \left(\lambda_k - \mathrm{MC_i}\right)\left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) \geq 0 \qquad \forall i > 0 \tag{1.97}$$

If we consider which technology is best placed to generate with a utilisation level of $u_1$, then:

$$\mathrm{FC_i} \geq \left(\lambda_1 - \mathrm{MC_i}\right)\mathrm{u_1} \qquad \forall i > 0 \tag{1.98}$$

This must hold for each technology at $u_1$, and must hold with equality for the marginal technology i*, assuming for the moment that there is only a single marginal technology. To be clear, we are specifically considering the cost recovery of technologies entering the merit order at this utilisation level, and not those technologies operating at higher utilisation levels that have additional cost recovery opportunities. Therefore, the energy price, $\lambda_1$, is the minimum price at which any technology can achieve cost recovery while operating at that utilisation level. We denote that optimal technology i*, so that the price is defined in terms of the cost parameters of i* :

$$\lambda_1 = \mathrm{MC_{i*}} + \frac{\mathrm{FC_{i*}}}{\mathrm{u_1}} \tag{1.99}$$

In conventional optimisations, this price is determined by the equilibrium investment condition to ensure that the marginal profitability of additional capacity is equal to the fixed cost of the technology,

mirroring the concept of uplift prices discussed in in Sherali (1982). Were generators to offer this as their "marginal" price, the spot market would discover the prices necessary to provide cost recovery. This would accord with a dynamic view of efficiency, however such an offer would be incompatible with perfect competition in the spot market. If such pricing were possible, then it could also be implemented to recoup investment mistakes, over-investment, or many other myriad reasons. Whatever the reason, the ability to charge such prices suggests some market power, and is therefore not price-taking behaviour

### 1.6.3 Optimisation with Piecewise Constant Load Classes.

In Section 1.6.2 the restrictions on generation functions are artificial, resulting from the LDC definition. In this section we consider a particular version of the LDC in which the actual LDC structure, as opposed to our representation of it, implicitly restricts generation functions even when generation functions are free to use any breakpoints in the [0,1] interval. This distinction is significant in terms of the second FTWE.

The second FTWE states that a Pareto-efficient allocation can be supported by a competitive equilibrium. If we do assume the LDC is piecewise constant, then we have a case where the true optimal solution of the general formulation, which is also a Pareto allocation, is not supported by the competitive equilibrium defined in the spot market. To be clear, while there is a competitive equilibrium capable of supporting that Pareto-efficient allocation of resources, it is not a competitive equilibrium that matches the circumstances being analysed. Specifically, when the LDC is piecewise constant, cost recovery does requires some apportionment of capacity costs to provide cost recovery, but the pre-determined nature of the spot market clearance mechanism cannot provide that.

The fundamental reason for this is that the optimal utilisation levels associated with the operation of the marginal technology in those load classes are implicitly determined by the structure of the LDC so even when utilisation levels corresponding to optimal trade-offs are included in the LDC definition, these are not selected. In this case, the optimal trade-off developed by the screening curve does not apply. As a result, optimal cost recovery is not attained and each technology will need some form of uplift for it to achieve cost recovery. To illustrate the issue, we present a reduced example of that used in Section 1.6.2. In this example, we assume the true LDC is piecewise constant with two load classes defined as:

| | | |
|---|---|---|
| Load (MW) | 80,000 | 68,000 |
| Maximum Utilisation | 0.1 | 1 |

There are two generation technologies plus a shortage technology with costs defined as:

| Technology | Fixed Cost | Variable Cost |
|---|---|---|
| Notional Shortage | 0 | 15000 |
| OCGT | 50 | 3500 |
| Geothermal | 1000 | 0 |

In terms of capacity, the single stage model has the following optimal solution:

| Technology | Capacity |
|---|---|
| Notional Shortage | 0 |
| OCGT | 12000 |
| Geothermal | 68000 |

The single stage formulation requires price uplift terms to achieve cost recovery. In this case, the OCGT technology must receive \$4,000 when marginal, to recover \$50 for each unit of installed capacity when operating with a utilisation of 0.1. The geothermal technology captures \$400 towards cost recovery while OCGT is marginal and must recoup \$600 while marginal, implying a price of \$666.67.

Whether or not cost recovery is achievable, depends on whether or not the spot market discovers those prices. The following condition defines the spot market price, which is actually a range when at a boundary point.

$$-\lambda_k + \mathrm{MC_i} + \varphi^{+}_{i,k} - \varphi^{-}_{i,k} \geq 0 \qquad (1.100)$$

In this example, we are on a boundary in both load classes as the capacity of generating technologies coincides with the load requirement of the load class. Piecewise constant LDC formulations will necessarily contain boundary cases, but in general there may also be intermediate load classes for which the boundary discussion does not apply, and for which the above price condition is not ambiguous. However, for those load classes where boundary conditions arise, and generation has exhausted all capacity of the marginal technology, the price is ambiguous. The prices required for cost recovery are available as solutions to the condition above so the spot market has the flexibility to deliver cost recovery if by some chance the correct price was selected. But, within the range for which the above condition is satisfied, there is no mechanism for determining these prices precisely. We note that the same issue arises whenever a boundary values is reached and this happens and this can also happen in a model with piecewise linear LDC's but the price that results in those cases is an instantaneous price, and does not span a wide utilisation range.

The nature of price flexibility at boundary points can be illustrated further by considering the marginal benefit of investment function. The marginal benefit of investment is a representation of the call option valuation of the plant. It is the additional profit that an additional unit of capacity would receive. In Figure 15, which is drawn for a more generic situation than our example, the marginal technology in each load class is initially fixed as capacity increases from zero. Eventually, marginal generating technologies change in individual load classes causing discrete price movement in the load class concerned. To be clear, as the marginal benefit of investment represents the weighted profitability of the technology, discrete reductions in the marginal benefit of investment result.

For example, if we consider the possibility of increasing geothermal capacity. The diagram depicts a more nuanced market with more load classes than our simple numerical example, but we might of increasing geothermal capacity. In the above example, this would occur when geothermal capacity reaches a level sufficient to change the marginal generating technology in a load class. In this simple example that level would be 68,000 at which point geothermal becomes the marginal generating technology. In addition, that level of capacity would also change the marginal generating technology in the other load class. In this example these occurred simultaneously, but in general they could occur separately, and either one would have the same effect on the marginal benefit function. In terms of the diagram, the additional steps reflect the possibility of additional load classes so that, for example if there were a load class with load of 74,000, then as soon as geothermal capacity reached 62,000, this load class would no longer have shortage as its marginal technology, creating a step in the function shown in Figure 15.

This adjustment is not problematic until we reach the level shown at A. At A, further investment is signalled as being profitable. Yet if it occurs a discrete drop in pricing will ensure that the technology will no longer be able to cover costs. This is shown at B, where consequently the market is signalling a decrease in capacity is desired.

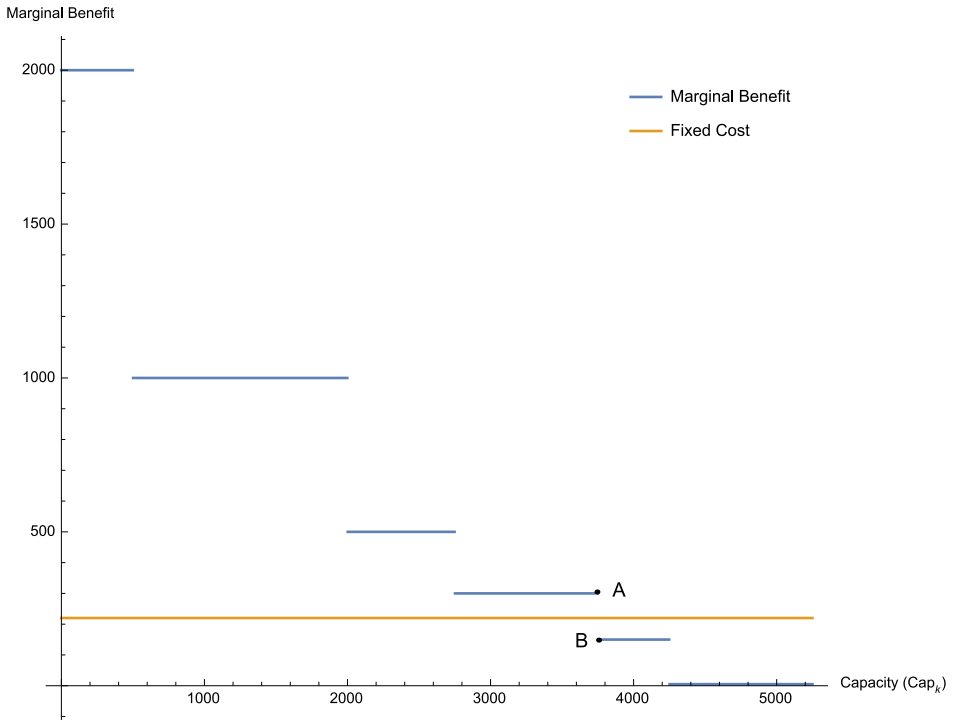



**Figure 15: Marginal Benefit of Investment: Conventional Optimisation**

As we have seen above, at the capacity level shared by A and B, at least one of the individual spot market prices applicable to a load class is at a boundary point. The corresponding price(s), which are the dual variables on which the marginal benefit of investment is based, are only defined as sub-gradients at these boundary points.

If we assume a pessimistic investor, we would assume the worst-case prices chosen from the sub-gradient range. If the situation was truly as modelled, one might expect those prices. If prices

were determined at the lower end of the feasible range then no technology would achieve cost recovery, and investors would not invest if they believed this was the payoff they would receive. If pricing resolved to these levels it is clear that, as a slight decrease in investment would benefit all participants, the solution is not Pareto optimal and therefore cannot be an competitive equilibrium.

But, for a given market in which the LDC was an approximation, the 'average' price in the spot market for each load class is not likely to be the pessimistic price. It will be between the most pessimistic price and the most optimistic one. However, when considering reality, it is important to realise that all investors will not likely perceive a piecewise constant LDC, as this is only a modelling convention. Rational investors will understand this, and consider that just as other investors not will perceive (the same or any) arbitrary restrictions on generation functions, other investors may not consider the LDC to have the same piecewise structure.

If prices were determined at the higher end of the feasible range, then all technologies would be profitable, and achieve more than cost recovery. In this case further investment is incentivised and so this is not an optimal solution. However, any further investment would instantly remove all flexibility in the condition above, as there would no longer be an alignment between capacity levels and the constant load level within load classes in which the newly invested technology operated.

The implication is that there is no consistent solution to the piecewise constant LDC implementation, even with a general form. The true nature of the LDC structure, rather than our representation of it, restricts the set of possible utilisation levels. For this reason, for the rest of the thesis, we will not consider piecewise linear LDC implementations.

## *1.7 Summary and Conclusions*

In this chapter we have reviewed the fundamentals of investment analysis in electricity markets. The assumptions of screening curve analysis are, in the main, acceptable in the context of an investment planning model predicated on perfect competition. The atomistic nature of perfect competition provides support for the implicit assumptions of continuous investment and the absence of economies of scale in generation, whether those are expressed in the form of the actual cost structure, or implied by minimum operating levels. The single node structure of screening curve analysis does eliminate the possibility of losses but these can be approximated by a priori modification of load levels that capture both the energy cost and the capacity implications of losses. Under perfect competition with free entry, particular locational issues resulting from network congestion could reasonably be assumed to be transient.

One of the primary benefits of screening curve analysis is the expression of optimal trade-offs, between generation technologies and more broadly generation and shortage, as defined by a VOLL. In the absence of complications, that analysis extends to a precise definition of the PDC, independent of the LDC. Screening curve analysis also enables a connection to be drawn between the returns of a particular technology and its option value. Other than the stochasticity imbedded in the LDC, the models in this chapter are deterministic. As a result, the concept of option value concept is somewhat redundant, however in more general circumstances involving stochasticity the same principles apply and the logic of option representation provides a basis for considering investment.

As complexity increased, the ability of screening curve analysis to satisfactorily incorporate such complexity generally decreased. This practical realisation has motivated widespread introduction of optimisation-based techniques to resolve investment and capacity planning problems. The problem of investment and generation is a two-stage problem although it can be formulated as an optimisation under perfect competition. That general optimisation formulation is presented in this chapter. As the lower level generation problem is a sub-gradient of the upper level generation problem, we are also able to define a single stage optimisation that represents the problem.

These formulations are conceptual and cannot be implemented as written. They contain no definition of the LDC, or any definition of the functional form of the generation function, other than the requirement that it be integrable over the utilisation range [0,1]. To implement the formulation requires specialising it, by defining the LDC and restricting the functional form of the generation function. These specialisations are conventional optimisation formulations.

By way of example we examine implementations of the conventional optimisation formulation using piecewise constant LDC's, piecewise linear LDC's as well as higher order piecewise representations. In the case of piecewise constant LDC's we note that within a load class a consistent representation of both energy and capacity requirements is elusive. The piecewise linear LDC formulation goes some way to resolving that issue by including two profiles, or basis functions in the load representation of a load class. This creates two prices, one for each profile, but the marginal technology is clear in each load class. Higher order forms offer improved fitting possibilities but their introduction results in additional basis functions, which the conventional optimisation formulation does not clear in accordance with merit order principles. For example, there can be a linear and a quadratic load profile generating at the same time, both with spare capacity.

In practice, the functional form of the generation function follows the functional form of the LDC, but this is not the problematic restriction. There is an implicit restriction on the breakpoints of the generation function: they must coincide with the breakpoints that define the LDC. The impact of this restriction is significant. In Section 1.6.2, we demonstrated that this restriction means the solution to the conventional optimisation formulation is not a solution to the general optimisation. We demonstrate this in two ways. Firstly, using an assumed true piecewise linear representation as an example we illustrate the objective function is sub-optimal relative to the actual optimal solution, even when the LDC is perfectly defined. Secondly, by illustrating the solution is not Pareto-efficient, it therefore cannot be optimal. Further, as the solution is not Pareto-optimal, then by the Fundamental Theorems of Welfare Economics, the solution cannot be a competitive equilibrium.

There are also a number of practical implications for conventional optimisation formulations that would be of concern to potential investors with rational expectations. Firstly, as the solution is not Pareto-efficient, it cannot be viewed as an equilibrium at all, competitive or otherwise from the perspective of investors. Investors will rationally expect further trading to occur. Secondly, to provide cost recovery, conventional optimisation formulations must in general add uplift payments to the marginal cost of the marginal technology. Rational investors will understand there is no basis for discovering these prices in the spot market clearing process. Furthermore, they will naturally ask the

question of what the PDC will look like given the prescribed capacity choices. Without uplift pricing there cannot be cost recovery.

The underlying cause of these anomalies is the hidden restriction on the structure of generation functions in conventional formulations. We know from Section 1.4 that basic screening curve analysis prescribes that generation functions have breakpoints corresponding to optimal technological trade-offs. As load increases, when one technology supplants another, that technology begins generating while the other technology maintains output at capacity. These breakpoints are necessary to define the interaction of adjacent technologies and optimal cost recovery through a perfectly competitively priced PDC.

In Section 1.6.3, we describe the important example special case of piecewise constant LDC formulations. As a direct result of the piecewise constant structure, the optimal trade-off between individual technologies is not respected. In this case, even if the generation were unrestricted and utilisation levels corresponding to optimal trade-offs were introduced to the LDC definition, the issue would remain as the generation functions are implicitly restricted by the incentives of the situation, over and above the representation of the LDC. Furthermore, we have a coincidence between the generation level in certain load classes and the level of capacity of all generating technologies. At such points, the market clearing price is ambiguous, and investors have no justification in assuming the price will gravitate to any particular level within the range defined. To be clear, the same ambiguity results in other formulations, but in those formulations they are only instantaneous prices, rather than prices applicable for the width of a load class.

To our knowledge, the implicit restrictions inherent in conventional optimisation formulations have not been described definitively. A lack of evidence to the contrary suggests that it is not widely understood that even when the LDC is perfectly represented, the solution to a conventional optimisation function will not represent either the optimal solution of the problem, or a competitive equilibrium. Noting this, and the nature of the irregularities in conventional optimisation formulations, and utilising the strong logical basis underpinning screening curve analysis, we now develop an approach that synthesise the two, and avoids the pitfalls of the conventional optimisation formulation.

# 2 ENDOGENOUS UTILISATION LEVELS

## *2.1 Introduction*

In the previous chapter we identified that while conventional optimisation formulations have many computational advantages and are readily extended, they produce solutions that are not:

- Optimal relative to the actual problem even when the LDC is precisely represented
- Representative of competitive equilibrium
- Internally consistent
- Suitable for sensitivity analysis.

As noted in Section 1.5.3, the actual problem is a two-stage problem. Conventional optimisation formulations produce prices that contain terms related to capacity cost recovery, although there is no such mechanism available in the spot-market clearing mechanism to provide discovery of such prices. We also know from Section 1.5.3 that there is an optimisation available that is equivalent to the two-stage problem formulation presented that is equivalent, so there is no apparent issue with the problem staging under perfect competition. Chapter 1 also included a discussion of screening curve analysis. In comparison with optimisation techniques, and the problems identified in the application of comparatively simpler optimisation techniques, screening curve analysis is shown to have some conceptual advantages, at least in simple problems:

- The optimal solution represents a perfectly competitive equilibrium that is not in conflict with the implications of the spot market clearing system.
- The utilisation levels of each technology, and their optimal marginal operating ranges are easily identified in almost all cases
- The equilibrium PDC is also readily, and precisely, identified.

Given the comparative advantages of optimisation and screening curve analysis, and the need to consider complementarity for the purpose of generalising future analysis to include risk, for example, in this chapter we seek to resolve the methods by developing a complementarity modelling framework that does identify the competitive equilibrium and/or optimal solution of the problem at hand, and will be relatively extensible when it comes to considering other issues.

Following a more direct comparison of screening curves and conventional optimisation formulations in Section 2.2, we commence the development of a consistent approach in Section 2.3. That process begins with a basic complementarity representation of a conventional investment formulation and a conventional spot market clearing formulation, each of which are then generalised to account for sub-periods. Utilising the intuition of screening curve analysis, we define the optimal trade-offs between technologies using a novel complementarity framework that recognises the bi-directional nature of trade-offs. This generalisation is done in a particular fashion in order to facilitate the calculation of sub-period trade-offs so that sub-period PDC's may be optimally defined. Having defined the full set of optimal technological trade-offs, we present two options for pruning the set of pairwise optimal trade-offs to define a smaller sub-set of only the critical trade-offs that define the screening curve lower envelope. The first method is simpler and computationally less onerous,

offering less than the maximum pruning of the utilisation set, whereas the second is significantly more complex and enables pruning of the utilisation set to the degree the modeller chooses.

Each endogenous utilisation level corresponds to an endogenous load level, and we present a set of complementarity conditions to calculate the corresponding load levels in the context of a piecewise linear LDC definition. Finally, we introduce complementarity conditions to order the endogenous utilisation levels and integrate these with the exogenous utilisation levels that define the LDC representation.

In totality, the complementarity formulation designed to achieve this is complex, so we consider other solution approaches including a nested approach and a decomposition approach. As the complementarity formulation is known to have multiple, albeit practically identical, solutions, the decomposition approach provides a useful background for the discussion of solution properties. Considering the complexity required to precisely define the solution, we make clear that the conventional approach, which is approximate even when data precisely matches reality, may be desirable on purely computational grounds. We temper that enthusiasm by noting that for problems that require complementarity formulations for structural reasons, the computational cost of significantly increasing the problem size will be significantly higher than when the underlying approximation is a convex optimisation.

## *2.2 Conventional Optimisation & Screening Curves*

In Chapter 1 we considered screening curves before addressing optimisation formulations in some detail. Consideration could be also be given to the adoption of a two stage modelling paradigm such as MPEC or EPEC which could clarify the nature of adjustment in optimisation models. However these approaches are not ultimately helpful in resolving the inconsistency as it is not a result of staging but a result of an implicit restrictions on generation functions and utilisation levels that follow from the definition of the load classes, load slices, or any other points used as a basis for developing interpolated solutions. Therefore, we restrict our focus to conventional optimisation methods and screening curves, as it is these two approaches that are most relevant.

Having identified some of the issues with conventional optimisation formulations, we consider the difference between conventional formulations and screening curves in order to devise an approach that resolves the issues uncovered and provides a basis for further analysis of the type envisaged at the commencement of this research. As noted earlier, from this point the discussion focusses on piecewise linear or higher form LDC representations, and specifically excludes the case of piecewise constant LDC's. Despite the relative complexity of modern approaches, when compared to screening curve analysis, it is the simplicity of the screening curve diagram, being a relatively elementary analysis, which avoids many of the deficiencies of the conventional optimisation model. We compare these two broad approaches in terms of utilisation levels, PDC definition, and the nature of equilibration in each.

### 2.2.1 Utilisation Levels

In the process of implementing a conventional optimisation formulation, we also introduce implicit restrictions on the solution as we use the LDC approximation not only to define the LDC used in the

model, but also the breakpoints of the generation functions. While the selection of utilisation levels is endogenous, the set from which they are selected from is not, as it is determined by the LDC definition.

In contrast, screening curve analysis does not accept utilisation levels as an input, instead producing them as an output. These, and the marginal operating range for each technology, are dependent on the relative cost structures of different technologies. Critically, at these endogenously determined utilisation levels, the cost of building and generating with the two adjoining technologies is equal, and therefore Pareto-optimal.

### 2.2.2 PDC Definition

As we have shown, in screening curve analysis, the optimal PDC is determined automatically by the optimal utilisation ranges. By definition cost recovery is achieved. We can see this by considering the shortage technology with zero fixed costs. It is only replaced when the next technology can cover its fixed costs. The progression of other technologies follows through optimal trade-offs, each recovering the incremental fixed costs as the optimal plant mix is defined.

From the basic logic of underlying optimal technological trade-offs, we know that for a conventional optimisation to produce the true optimal solution, it must necessarily include in the LDC representation the utilisation levels that correspond to technological trade-offs, or it will not be possible to replicate the true optimal PDC, where such a PDC exists.

Where these utilisation levels are not included, the conventional optimisation formulation produces a PDC which incorporates the effects of artificially imposing restrictions on the form of generation functions. We are faced with prices that include cost recovery components that are not discoverable by the spot market clearing process. Taking the example of the ultimate marginal technology, we have the following decomposition of the price showing a cost recovery component:

$$\lambda_1 = \mathrm{MC}_{i^*} + \frac{\mathrm{FC}_{i^*}}{u_1} \tag{2.1}$$

As described in Section 1.6.2, the requirement for an uplift term exists more generally. It is worth noting the difference between a competitive equilibrium that responds to incentives, and a central planner in this context. The existence of a sub-gradient suffices for a central planner, who can then choose a capacity level that maximises the gains of installing capacity and, if they so choose, a set a price that results in cost recovery. Although in slightly different contexts, Rothkopf et al (2004), or Bjorndal & Jornsten (Bjorndal & Jornsten, 2008) show that a price is available that supports an equilibrium in the presence of non-convexities, and this price is composed of a commodity price and an uplift. As with Sherali et al (1982), they have developed pricing mechanisms that support the optimal primal result, which in this case means that cost recovery requirements define pricing.

But in a decomposed competitive system, it is the other way around. Pricing must support cost recovery, and pricing is determined by a market clearing process in which the mechanism of perfect competition determines pricing. Because there is no mechanism in the problem structure to allow for prices in the spot market to provide cost recovery explicitly, this PDC includes prices that will not eventuate, leading to the rather decayed PDC in Section 1.6.2 that results from market clearances using the capacity implied by the conventional optimisation.

Ideally, we would like to correct the PDC generated by the conventional optimisation by removing the unnecessary restrictions that the conventional optimisation approach places on it. This would drive the fixed cost recovery components to zero, as they are in screening curve analysis. We do that in the remaining sections and Chapters of this thesis. But if we proceed with the conventional optimisation formulation, we can mitigate the degree of PDC misrepresentation.

As shown in the example of Section 1.6.2, even when the objective function of the conventional optimisation is relatively close to the true optimal solution, the plant mix is not. The inaccuracy of the solution is directly related to the inaccuracy of the PDC as these mutually determine each other. To optimally define the PDC, we must include utilisation levels from screening curve analysis that reflect optimal technological trade-offs. A priori knowledge of these utilisation levels implies knowledge of the optimal solution, so simply requiring the addition of some utilisation levels to the formulation is unhelpful.

Nevertheless, the solution accuracy can be improved by selecting more appropriate utilisation levels. To increase the accuracy of the PDC, modellers often appear to make a priori estimates of utilisation levels that are close to the optimal levels that might be expected. For example, in Ehrenmann & Smeers (2011), the utilisation levels used to define the LDC coincided closely with the optimal trade-offs between technologies in that study. This is approach is common and reasonable but in many instances simply using a utilisation level close to the actual optimal trade-off is not as helpful as it might seem. For example, in the example in Section 1.6.2, the optimal shortage frequency is 0.00435 (5dp). Should the modeller have guessed and implemented a utilisation level of 0.005, the conventional optimisation would have concluded that there would be no shortage.

Another approach is to introduce a significant number of utilisation levels. Instead of attempting to define the utilisation levels required, this approach includes an extremely granular representation of the LDC in an attempt to capture or approximate the critical levels that are required. Revisiting the first order optimality conditions for generation, we see that the price is defined as the marginal cost of the marginal technology.

$$\frac{\partial z}{\partial GEN_{i,k}} = -\lambda_k + \mathrm{MC_i} + \varphi^{+}_{i,k} - \varphi^{-}_{i,k} \geq 0 \qquad \forall i,k < K \qquad (2.2)$$

By increasing the number of load classes, through the inclusion of additional utilisation levels, the proportion of load classes in which the price is misrepresented decreases, and the width of the load classes in which the price is misrepresented also decreases. Accordingly, the accuracy of the PDC and the investment decisions based on it increases. Whereas the conventional approach effectively discretises the utilisation range, as we increase the number of the utilisation levels that are considered then, in the limit, we approach a continuous range which we can be assured contains the critical utilisation levels required.

There are some downsides to increasing the granularity of the problem. Firstly, it is difficult to know what level of granularity is required to achieve a given accuracy standard. Furthermore, increasing granularity comes at a computational cost, although in the context of fast convex optimisation solvers that cost is likely to be small enough to be tolerable on simple problems. However, the cost of increasing granularity in a complementarity formulation is computationally much

higher. In those cases an intelligent method that intelligently defines the problem by determining the required utilisation levels may be more effective.

### 2.2.3 Investment Equilibration

Equilibration of investment occurs through the marginal benefit of investment function, from which a capacity level will be chosen that equates the marginal benefit of investment/capacity with the fixed cost of capacity. For convenience we repeat the relevant investment condition here:

$$\mathrm{FC_i} - \sum_k \varphi^+_{i,k} - \chi^-_i \geq 0 \tag{2.3}$$

As can be seen from the equation, the adjustment is slightly more nuanced than stated above. There are bounds, including non-negativity, that can prevent the equilibration but we do not concern ourselves with those cases until later chapters.

The structure of the marginal benefit of investment function that results when implementing a conventional optimisation approach mirrors the structure of the function when the LDC is piecewise linear. That is, it is stepwise constant, with steps corresponding to capacity levels for which the spot market price is constant in load classes. Although the underlying reason for the implicit restriction of utilisation levels is different, the fact that utilisation levels are restricted determines the structure of the marginal benefit of investment function in both cases. We repeat this in Figure 16

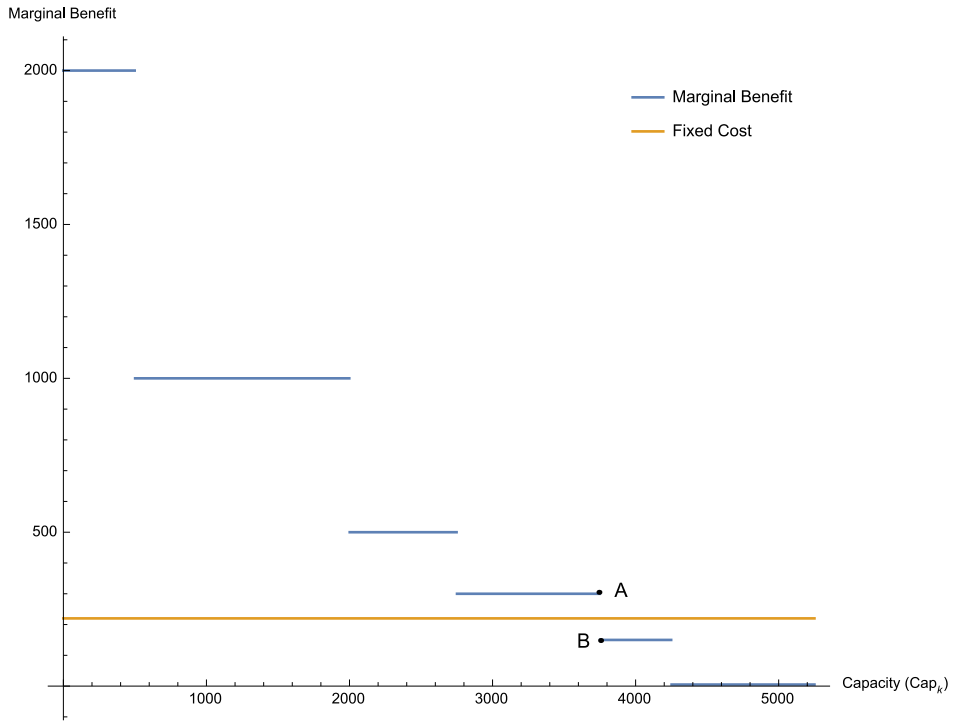



**Figure 16: Marginal Benefit of Investment Function: Conventional Optimisation**

That adjustment process is discrete which is unrealistic. An illustrative example involves considering a change in the cost structure. As the cost structure of a generation technology there is initially no change in the utilisation of the technology as the PDC is constant. Instead of changes in the equilibrium usage of the technology and its neighbours, the slack in cost recovery is taken up by

changes in the cost recovery portion of the optimal price. The implication is that changes in cost are initially borne by the market, which is inconsistent within the specified perfectly competitive structure of the spot market. Eventually there is a change in the optimal plant mix, which will be discrete, completing an adjustment of technological utilisation that is characterised by general invariance in response to cost changes, interspersed with sporadic and discrete changes in utilisation. We know from screening curve analysis that within the available utilisation range [0,1], continuous adjustment of cost increases result in continuous adjustment of utilisation.

If we suppose the modeller is interested in shortages or the revenue that comes from shortages, it is easy to see this behaviour is not desirable. The implication here is that the dynamic adjustment in the conventional optimisation formulation is one-dimensional and discrete. For example, a modeller examining the sensitivity of shortage to VOLL estimates would conclude that above a certain limiting value for VOLL, there is no shortage implied, while below that value shortage will occur in discrete tranches. It does not represent the two dimensional adjustment of equilibrium capacity and utilisation that would happen if, for example, variable costs were to drift upwards.

In screening curve analysis utilisation levels are endogenous and the equilibration of investment is significantly different. This adjustment assists the equilibration by providing the necessary degree of freedom to equate to the marginal benefit of capacity with the fixed cost of capacity. Figure 17 shows an example of such adjustment in terms of the PDC. In this example we consider the addition of capacity of a mid-merit order generation technology, in this case CCGT. As adjustment occurs, the PDC to the left of the technology of interest is moved sideways, and in this case the additional CCGT capacity entirely crowds out shortage and partially crowds out the OCGT technology. With fixed utilisation levels, the adjustment process would have resulted in no change in the PDC until the extra CCGT capacity caused a change in the marginal generating technology in one of the load classes.

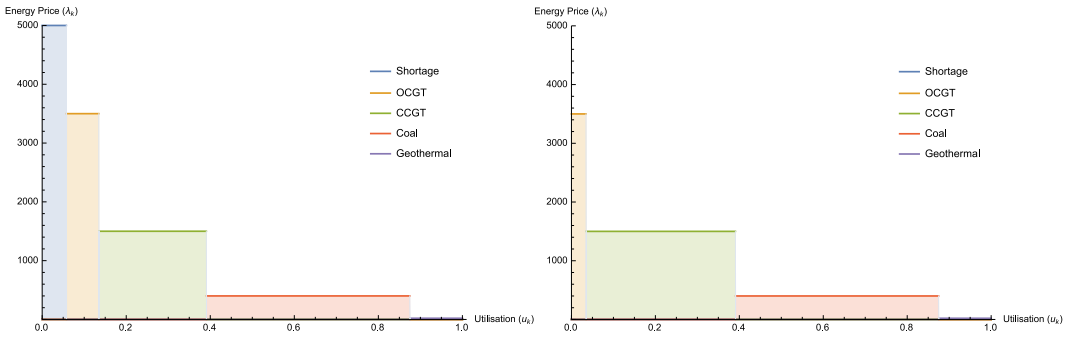



**Figure 17: PDC Adjustment**

To understand the mechanics of endogenous utilisation levels, we consider the marginal-benefit of investment function when equilibrium utilisation levels are endogenous. There are two separate relationships that determine the specific form of the marginal benefit of investment in this framework: the rate at which the marginal benefit of investment decreases with respect to changes to the

technology ultimately being supplanted, and the rate at which the marginal benefit of investment decreases with respect to utilisation of the capacity, which itself is an LDC dependent function of the level of capacity.

Before we consider the first issue, we elaborate on the relationship between changing capacity and the utilisation of technologies higher in the merit order. The rate at which utilisation ranges adjust is determined by the slope of the LDC at all levels above those that are served by the technology under analysis. Capacity increases directly induce changes in the utilisation of technologies higher in the merit order. These changes are relatively smaller in ranges in which the LDC is steeper whereas, where the LDC is flatter, increases in capacity induce relatively larger changes in utilisation of technologies higher in the merit order. Therefore, as this effect cascades upwards, it is not necessarily the case that the marginal operating range of a technology will be monotonically increasing or decreasing. For that to be the case, the slopes of each successive piecewise segment of the LDC would also have to evolve monotonically.

If we consider the simpler case of a linear LDC representation, then capacity adjustment of technology i will, with the exception of the ultimate technology, leave the width of marginal operating ranges of each technology higher in the merit order unchanged but will shift them consistently, according to the following rate:

$$\frac{\partial u}{\partial CAP_i} = \frac{\mathrm{L_k} - \mathrm{L_{k+1}}}{\mathrm{u_k} - \mathrm{u_{k+1}}} < 0 \tag{2.4}$$

By virtue of its utilisation range being bounded by zero, the generator of last resort, or notional shortage technology if that is the ultimate technology in use, is the exception. Rather than maintaining a constant marginal operating range, the marginal operating range of these technologies is reduced by the same fraction of time as the technology being introduced will operate. These technologies are effectively crowded out.

Geometrically, the marginal benefit of an additional unit of capacity is given by the option value defined the PDC and the marginal cost, or effective strike price, of that technology. As capacity increases, the utilisation of technologies higher in the merit order is reduced with the consequence being the marginal benefit of the introduced capacity is reduced. Within each capacity range characterised by a single marginal technology, the rate of adjustment of the marginal benefit of capacity is given by:

$$\frac{\partial MB_i}{\partial CAP_i} = \varphi_{i,1} \frac{\partial u_1}{\partial CAP_i} = \varphi_{i,1} \frac{\mathrm{L_0} - \mathrm{L_1}}{\mathrm{u_0} - \mathrm{u_1}} < 0 \tag{2.5}$$

The rate of that reduction depends on which technology or shortage cost represents the technology of last resort, as it is this technology that is ultimately supplanted by increasing capacity. The nature of the merit order process dictates that as capacity is increased, the marginal benefit of investment will decrease at an increasingly slower rate as initially high cost, and high price, technologies are eliminated from the dispatch, and so on down the merit order to lower cost, and price, technologies, whose impact on the option value defined by the PDC is lower.

In combination, these adjustments result in the piecewise linear marginal benefit function shown in Figure 18, in which each linear segment exists over a range of capacity levels, each of which correspond to a particular marginal technology in the process of being crowded out.

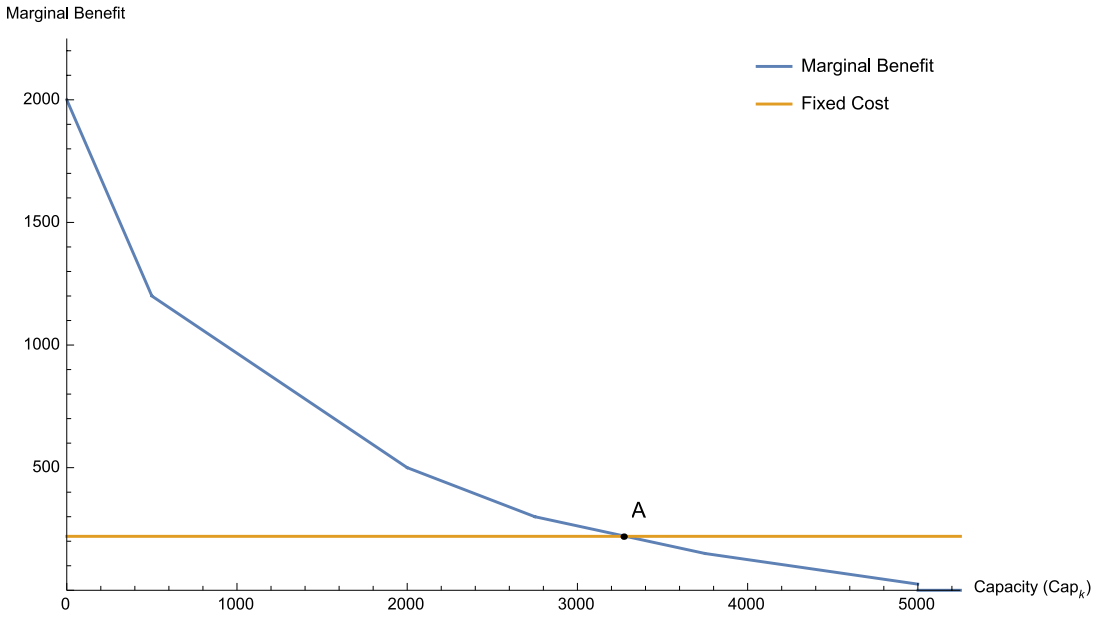



**Figure 18: Marginal Benefit of Investment with Endogenous Utilisation Levels**

Ceterus paribus, the equilibrium level of capacity is defined as shown at A, where the marginal benefit function is equal to the fixed cost component of the technology. The functional form in Figure 18 applies when the LDC is linear. But where the LDC slope varies from one piecewise segment to another, the rate of adjustment in the utilisation level with respect to capacity also changes. If this case, as capacity is introduced, then even while there may be no change in the technology being supplanted, the rate of decrease in the marginal benefit function may increase or decrease, depending on whether the LDC is becoming flatter or steeper. This will cause the slope of the marginal benefit function to adjust in a non-monotonic fashion. While that is a possibility, the adjustment of the PDC makes clear that the marginal benefit function itself retains its monotonically decreasing structure as the option value must shrink monotonically when additional capacity is added.

## *2.3 Complementarity Formulation*

### 2.3.1 A Consistent Approach

In this section we show that, by choosing a particular set of utilisation levels, we can synthesise screening curve and optimisation analysis to ensure modelled clearances will be consistent with actual spot market clearances and the precepts of perfect competition. This enables us to precisely solve the problem.

Screening curve analysis identifies the utilisation levels that correspond to optimal trade-offs between individual technologies, and these are the utilisation levels we require. In general, these utilisation levels will not be included amongst those chosen to represent the LDC. However, when

utilisation levels corresponding to optimal trade-offs are included in the analysis, the prices that would be generated using a simultaneous approach do coincide with those required for fixed cost recovery, in addition to being consistent with perfect competition.

To see this, consider the optimal trade-off between two technologies, A and B, at $u_k$ where technology B is the lower marginal cost technology, and technology A is the higher marginal cost technology. As technology B is at full capacity at $u_k$, if the market price is to be consistent with perfect competition we require $\lambda_k = MC_A$. Equally, there is a price that is consistent with the cost recovery of technology B, $\lambda_k = MC_B + \varphi^+_{B,k}$. As we have shown in Section 1.6.2 when using exogenous utilisation levels, these prices will not generally coincide. However, they will coincide at $u_k$, if and only if $u_k$ is defined by the optimal trade-off between technologies A and B expressed below:

$$FC_A - FC_B - (MC_B - MC_A)u_k = 0 \tag{2.6}$$

Substituting the equilibrium cost recovery conditions for built technologies into (2.6) we have :

$$\sum_{j=0}^{k-1} \varphi^+_{A,j+1}\left(u_{j+1} - u_j\right) - \sum_{j=0}^{k-1} \varphi^+_{B,j+1}\left(u_{j+1} - u_j\right) - \left(MC_B - MC_A\right)u_k = 0 \tag{2.7}$$

Re-writing the final $u_k$ in (2.7) as the sum of individual load classes, and grouping terms gives:

$$\sum_{j=0}^{k-1}\left(u_{j+1} - u_j\right)\left(\varphi^+_{A,j} - \varphi^+_{B,j} + MC_A - MC_B\right) = 0 \tag{2.8}$$

But, as $u_{j+1} - u_j > 0$ and $\lambda_{j<k} = MC_i + \varphi^+_{i,j} \quad \forall i, j < k$ we have $\varphi^+_{A,j} - \varphi^+_{B,j} + MC_A - MC_B = 0 \quad \forall j < k$, implying:

$$\varphi^+_{A,k} - \varphi^+_{B,k} + MC_A - MC_B = 0 \tag{2.9}$$

From the market clearing condition, we have $\varphi^+_{A,k} = 0$, so that:

$$\lambda_k = MC_B + \varphi^+_{B,k} = MC_A \tag{2.10}$$

The market-clearing price is equal to the marginal cost of the marginal generator, A, which is consistent with perfect competition. In addition, this is the price implied by the investment condition, which guarantees cost recovery for all built technologies at $u_k$.

Having established the rationale for including utilisation levels that correspond to optimal trade-offs, we turn our attention to the introduction of these particular utilisation levels. Given the longer term goal of the research program embarked upon, and the flexibility of complementarity to describe in a uniform fashion optimisation, algorithmic and logical features, we adopt a modified complementarity approach, building up the complementarity restrictions from their economic foundations. It is possible that other formulations, such as non-linear programming might suffice but,

given the future applications of the framework, this would likely be unhelpful and eventually require a complementarity framework also. Furthermore, it is worth noting that, in many instances, the formulation type is only a notational difference given that current solvers may significantly reformulate problems to adapt particular problems to existing solution strategies. In any case, the logic underpinning those approaches will have to confront the same issues that we describe.

### 2.3.2 Complementarity & Optimisation

Complementarity problems and optimisation problems are closely related. One major source of complementarity conditions is the KKT conditions of an optimisation problem. Indeed, in this thesis a number of basic investment models provide a subset of complementarity conditions to larger formulations. In addition, there are other optimisations such as ordering or ranking optimisations that provide complementarity conditions also. There are also a significant number of complementarity conditions that do not arise from optimisation throughout the thesis. The advantage of a complementarity formulation relative to an optimisation formulation is that many sets of complementarity conditions can be considered jointly. For example, in the gaming literature, complementarity formulations are used to combine KKT conditions formed from the individual optimisation problems of agents to define an equilibrium. In electricity markets, an auction is often used for price discovery and the KKT conditions corresponding to the optimisation guiding that auction are included in this thesis. From a technical standpoint, one relative strength of complementarity is the ability to combine different modelling paradigms in a unified framework.

Forming a Lagrangian comprising the objective function and penalty terms that reference constraints, multiplied by the corresponding dual variables is the first step in discovering the KKT conditions of an optimisation problem. From this point, first order derivatives are formed with respect to both primal and dual variables. Complementarity slackness conditions require the product of each derivative with the associated variable to be zero. In combination, these conditions define the KKT conditions for the original optimisation problem.

For the KKT conditions to be meaningful, so that the solution of the KKT conditions also corresponds to the optimal solution of the original optimisation problem, certain constraint qualifications must be satisfied at the optimal solution. There are many constraint qualifications available but, in the case of the conventional optimisation formulations described in the thesis, a simple constraint qualification LICQ (Linearly Independent Constraint Qualification) suffices. Where constraint qualification is not satisfied, there may be solutions to the KKT conditions which are not the solution to the original optimisation, and therefore would not be Pareto optimal or correspond to the competitive equilibrium we seek.

In this chapter, complementarity conditions corresponding to the inconsistent optimisation example in Chapter 1 are used. This appears inconsistent as, given constraint qualification is satisfied, the solution of each will be identical. However, the complementarity conditions used in Chapter 2 do not exist in isolation. There are a significant number of other complementarity conditions that have the practical effect of optimising the problem representation, and generating a consistent solution. These optimal utilisation levels appear in the KKT conditions used in the complementarity formulation as

well as in the KKT conditions of the conventional optimisation, but it is only in the former that they are variables. This will be illustrated further in Section 2.8.

### 2.3.3 Market Clearance

We begin by considering market outcomes with a fixed capital stock of each technology, $\mathrm{CAP_i}$, including a notional shortage technology i=0. Taking the relevant constraints and portion of the objective function from the formulation in Section 1.5.3, we can formulate the market clearance problem at a utilisation level $\mathrm{u_k}$ as follows:

$$\underset{GEN_{i,k}}{Minimise} \sum_i \mathrm{MC_i} GEN_{i,k} \tag{2.11}$$

$$\sum_i GEN_{i,k} - \mathrm{L_k} = 0 \quad : \lambda_k \qquad \forall k \tag{2.12}$$

$$\mathrm{CAP}_i - GEN_{i,k} \geq 0 \quad : \varphi^+_{i,k} \qquad \forall i > 0, k \tag{2.13}$$

$$GEN_{i,k} \geq 0 \quad : \varphi^-_{i,k} \qquad \forall i, k \tag{2.14}$$

Here we wish to draw the reader's attention to the dual variables in this formulation. To be clear, from this point on in this thesis, these have been redefined to apply to the context of a single market clearance for a single utilisation level. They no longer correspond to a load class (a range of utilisation levels) and have no load class or other scaling built in to their valuation. The behaviour of these variables in intervening periods is addressed later.

Feasibility of the above problem is guaranteed by the presence of a notional shortage technology. Provided the marginal cost values are distinct, and the load level does not precisely coincide with the sum of a subset of the available fixed capacities, then the power price, $\lambda_k$, may only assume a value from the full set of technological marginal costs, including shortage costs, and the problem has a unique solution. Recognising that an instance of this market clearance problem exists at each utilisation level, $u_k$, we present an equivalent complementarity formulation containing the equilibrium conditions for all instances of the problem above:

$$-\lambda_k + \mathrm{MC_i} + \varphi^+_{i,k} \geq 0 \quad \perp \quad GEN_{i,k} \geq 0 \qquad \forall i, k \tag{2.15}$$

$$\sum_i GEN_{i,k} - \mathrm{L_k} = 0 \quad \perp \quad \lambda_k \; free \qquad \forall k \tag{2.16}$$

$$\mathrm{CAP_i} - GEN_{i,k} \geq 0 \quad \perp \quad \varphi^+_{i,k} \geq 0 \qquad \forall i > 0, k \tag{2.17}$$

In conjunction with (2.16), which requires load be served, and (2.17), which describes maximum generation levels, (2.15) defines a market clearance with fixed capital stocks by relating the market price, $\lambda_k$, to the marginal cost of each technology at each utilisation level, $\mathrm{u_k}$. When technology i is marginal with spare capacity the market price is equal to the marginal cost of technology i. When

technology i is infra marginal, $\varphi^{+}_{i,k}$ reflects the profitability or rate of cost recovery of technology i at $u_k$. When technology i is not producing then its marginal cost exceeds the market price. We note that under these circumstances, a technology may not be both marginal and at full capacity, as there is no scope for increasing capacity, and supplying additional output at that marginal cost. The assumption of inelastic demand implies that, aside from coincidental cases, only the supply-side can be marginal and further illustrates that in the short-run, in which there is no opportunity to adjust capacity, then all market prices must be from the set of technological marginal costs.

### 2.3.4 Incorporating Investment

To correctly represent investment incentives, we first must understand the particular problem we have posed. In a traditional load class based formulation with a piecewise constant LDC, the price identified is applicable across the range of utilisation levels that define the load class. In effect the interpolation of the system occurs in the primal formulation and is achieved by dividing the LDC into load classes, each of which are defined by an interpolation of load. In this point-based formulation that is not the case. Instead, we assess market pricing at selected load levels and interpolate the behaviour of price, and therefore profitability between each of these points to develop the PDC. This amounts to interpolation in what would traditionally be considered the dual.

We must select a method for interpolating earnings that is consistent with a piecewise linear or, more generally, a non-piecewise constant LDC. In our formulation only a single technology will be marginal. Where technology A is that marginal technology at $u_k$, it will serve the incremental load of up to $L_k - L_{k+1}$ with partial use of its capacity across the utilisation range $\{u_k, u_{k+1}\}$. In the interior of that range, that technology will be price setting and the profitability of technology A will be $\varphi^{+}_{a,k+1} = 0$.

There remain several options for defining the equilibrium relationship between fixed costs and profits. Previously, the relationship was expressed in terms of dual variables, which in that formulation corresponded to the profitability or cost recovery flowing from a unit of generation across a load class. In this complementarity formulation, the dual is scaled differently, and $\varphi^{+}_{i,k}$ refers to the profitability of generation by technology i at a utilisation level, $u_k$. If we take an operational view and focus on total profitability and costs, a natural equilibrium relationship might be:

$$FC_i CAP_i - \sum_{k<K} \varphi^{+}_{i,k+1} \left( \frac{GEN_{i,k+1} + GEN_{i,k}}{2} \right) (u_{k+1} - u_k) \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{2.18}$$

Put simply, for technologies that are built, the cost of capital must be precisely covered by the total profit from generation, whereas when fixed costs exceed the total profits available, those technologies are not built. The constraint could also be expressed on a more traditional per unit of capacity basis. For technologies that are built this yields the following, where the average generation is normalised by $CAP_i > 0$:

$$\mathrm{FC_i} - \sum_{k<K} \varphi^{+}_{i,k+1} \frac{\left(GEN_{i,k+1} + GEN_{i,k}\right)}{2CAP_i} \left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) = 0 \qquad \forall i > 0 \qquad (2.19)$$

Conditions (2.18) and (2.19) can become mathematically problematic when capacity and generation are zero, so we are unable to rely on a relationship between total costs and total profitability. In equilibrium, no investment opportunities yielding positive profits should remain unexercised, but (2.18) and (2.19) fail to enforce this standard as they are backward looking conditions, that enforce an ex-post requirement that investors make at worst zero profit, thereby admitting the possibility of doing nothing, even though returns may be sufficient to justify investment. The underlying economic intuition that should motivate this constraint is forward, not backward, looking. We require a further simplification that does not directly depend on capacity and generation choices and relies only on system prices, and the distribution of profits that they imply for prospective investors in a particular technology.

As we are only focussed on a perfectly competitive spot market, with no other use for each technology and endogenous utilisation levels, such a simplification is available and it arises from the call option interpretation of the value of capacity. In a perfectly competitive market with no additional revenue sources available to the technology, whenever technology i is marginal at $\mathrm{u_k}$, its profitability, $\varphi^{+}_{i,k} = 0$. Therefore, its fixed costs must be recovered entirely when the technology is infra marginal. We can ignore the need to calculate average generation across those load classes in which capacity is partially utilised and instead only consider those load classes in which $CAP_i = GEN_{i,k+1} = GEN_{i,k}$. Under these conditions (2.19) simplifies to the form below:

$$\mathrm{FC_i} - \sum_{k<K} \varphi^{+}_{i,k+1} \left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \qquad (2.20)$$

This aligns with the option valuation of thermal plant that was discussed earlier, where $\varphi^{+}_{i,k}$ is the profitability, bounded by zero from below, that arises from operation in a given utilisation range defined by $\mathrm{u_{k+1}} - \mathrm{u_k}$. Interestingly, the form of the LDC is not important at this stage, provided utilisation levels are chosen appropriately. Where they are not, the simplification above is incorrect, as prices will include cost recovery components, resulting in profitable operations while being the marginal technology. In this case, the definition of total profitability is incorrect and would need to account for the functional form of the LDC representation.

Unlike the case with fixed capacity levels, when we consider investment, an additional response is available, and in general a particular technology can be both marginal and at full capacity at particular load levels. This provides an additional degree of freedom in (2.15) so that technological profitability and market prices are free within a certain range provided they operate in unison. For all technologies other than the notional shortage technology, this degree of freedom is eliminated by the additional complementarity condition relating to investment (2.20), which we are happy to use in the anticipation that future developments will deliver an optimal set of utilisation levels. Where the shortage technology is concerned, (2.20) does not apply and the system of complementarity equations

retains that unwelcome degree of freedom. That degree of freedom arises from the requirement to define a price at $u_0$, without that price being disciplined by entry through the investment criteria specified above. Mathematically, it follows from (2.15) that both $\lambda_0$ and $\varphi^+_{0,0}$ are free, provided they move in unison.

For the sake of future developments and applications, we need to ensure that these variables, and in particular, $\lambda_0$, are uniquely determined. One possible solution is to set the capacity of shortage to be higher than the maximum load, thereby ensuring that $\varphi^+_{0,0} = 0$, which in turn requires $\lambda_0 = \mathrm{MC}_0$ in times of shortage. While mathematically effective, the concept of "shortage capacity" is inappropriate, at least in this context. Rather than apply that mathematical workaround, we limit the range of technologies to which the complementarity conditions relating to capacity and investment apply to technologies $i > 0$. This distinction between shortage and actual generation technologies is conceptually consistent with the fact that shortages are not limited by capacity restrictions or an investment/market entry conditions.

### 2.3.5 Basic Investment Model

Market Equilibrium Conditions

$$-\lambda_k + \mathrm{MC_i} + \varphi^+_{i,k}\big|_{i>0} \geq 0 \quad \perp \quad GEN_i \geq 0 \qquad \forall i,k \tag{2.21}$$

$$\sum_i GEN_i - \mathrm{L_k} = 0 \quad \perp \quad \lambda_k \; free \qquad \forall k \tag{2.22}$$

$$CAP_\mathrm{i} - GEN_{i,k} \geq 0 \quad \perp \quad \varphi^+_{i,k} \geq 0 \qquad \forall i > 0, k \tag{2.23}$$

Investment Equilibrium Conditions

$$\mathrm{FC_i} - \sum_{k<K} \varphi^+_{i,k+1}\left(\mathrm{u_{k+1}} - \mathrm{u_k}\right) \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{2.24}$$

To summarise, market clearances generate a unique set of prices that define the price duration curve and guide investment decisions that feed back into the market clearance process as capacity levels. Because we are only considering a single period, the capacity requirement for each technology can be expressed in terms that are precisely defined by the optimal trade-off directly. In the following section we shall see that, in general, the capacity choice is a compromise based on the market outcomes from a variety of situations.

Aside from some basic scaling issues, these complementarity conditions represent the optimality conditions to the problem in Section 1.6.1. As that formulation is a LP, we have, barring the possibility of coincidence, a unique solution. We can also view the problem economically, using induction of the price determination argument. Apart from assisting our economic understanding of standard optimisation models in this field, the induction argument is constructive and suggests an alternative solution approach, which we discuss later in Section 2.7.

That problem contemplates a piecewise constant LDC. The complementarity conditions (2.21) to (2.24) could also contemplate a piecewise constant LDC model. In fact, where there are no further complicating factors, these complementarity conditions are suitable for any well defined LDC representation provided endogenous utilisation levels are chosen to define optimal marginal operating ranges. When endogenous utilisation levels are used, each technology is only profitable when operating at full capacity and so the intricacies of earnings while following a particular load profile corresponding to a basis function of the LDC approximation are given zero weighting. Conversely, whenever utilisation levels are not chosen appropriately, then the definition of plant profitability must necessarily be formulated to reflect the various possible load profiles that are feasible according to the LDC representation.

To be clear, while the complementarity formulation above can be written as an equivalent linear programming, this is only a subset of the conditions required to resolve the issues identified thus far. So far our exposition has been limited to a discussion of the behaviour of the complementarity conditions under circumstance in which utilisation levels have/have not been chosen endogenously. We need to add complementarity constraints to assist the setting of these variables. It is therefore meaningless to contemplate solution of the problem as a conventional optimisation as that approach would only be valid when paired with optimally chosen utilisation levels and, other than in the most trivial cases, those utilisation levels can only be determined by solving the problem itself.

### 2.3.6 Sub-Periods & Investment Decisions

The LDC typically represents the combination of several different processes that are then lost in the aggregation process. Provided that the underlying processes do not generate any other impacts on the system that would be correlated with load, this approach suffices. However, it is unlikely this is the case, and this creates a need to decompose the LDC further, in order to reflect more accurately those correlations more accurately. For example, the LDC typically has a strong seasonal structure, reflecting shifting power consumption as seasons and temperatures vary. It might also be the case that fuel prices and plant operations differ seasonally, as the role of intermittent and other renewable technologies varies across seasons also. The resulting correlation between load and fuel cost variations threatens the legitimacy of the summation of seasonal LDC's into an annual LDC. This example provides an example of the motivation behind the development of sub-periods in our model.

We now develop the formulation to incorporate sub-periods. Other than accommodating the potential for the type of correlations already discussed, the introduction of sub-periods at this early stage clarifies the different paradigm involved in optimal trade-offs with a single LDC as compared to optimal trade-offs over multiple LDC's. As we are operating in the context of investment decisions, our underlying goal remains the determination of the equilibrium plant mix through accurate and consistent development of the PDC, which underpins investment decisions. Before proceeding, we note that the sub-periods referred to need not be contiguous as in our example, and other forms of decomposition, such as by time of day, are equally valid where there are underlying variations and correlations that are best understood in those terms.

The sub-period/time dimension is indexed by t={1,…T}, and each sub-period occupies a fraction of the year $w_t$, where $0 < w_t \leq 1$. Each sub-period has its own LDC, defined by $\{L_{k,t}, u_{k,t}\} \forall k,t$. In addition to each sub-period having its own LDC, we allow marginal costs to vary between sub-periods, so that these are defined by $MC_{i,t}$. As capacity is assumed to be inflexible across the period that defines the full LDC, it must also be inflexible across a subset of that time period. For each technology, the decision of how much capacity to install is based on the aggregate returns available in each sub-period. Incorporating sub-periods and recognising the requirement for a single capacity choice, the market equilibrium conditions become:

$$-\lambda_{k,t} + MC_{i,t} + \varphi^{+}_{i,k,t}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,k,t} \geq 0 \qquad \forall i,k,t \tag{2.25}$$

$$\sum_i GEN_{i,k,t} - L_{k,t} = 0 \quad \perp \quad \lambda_{k,t} \; free \qquad \forall k,t \tag{2.26}$$

$$CAP_i - GEN_{i,k,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,k,t} \geq 0 \qquad \forall i > 0,k,t \tag{2.27}$$

Weighting sub-period returns by $w_t$ gives the new investment condition:

$$FC_i - \sum_t w_t \sum_{k<K} \varphi^{+}_{i,k+1,t} \left(u_{k+1,t} - u_{k,t}\right) \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{2.28}$$

While this formulation is sufficient and represents the basic formulation in complementarity form, it is more helpful to adopt a different formulation that explicitly defines sub-period performance. We begin by considering the optimal capacity of each technology as if capacity were perfectly flexible between sub-periods. We introduce the following complementarity conditions:

$$\chi_{i,t} - \sum_{k<K} \varphi^{+}_{i,k+1,t} \left(u_{k+1,t} - u_{k,t}\right) = 0 \quad \perp \quad CAP_{i,t} \; free \qquad \forall i,t \tag{2.29}$$

$$CAP_{i,t} - GEN_{i,k,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,k,t} \geq 0 \qquad \forall i > 0,k,t \tag{2.30}$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0,t \tag{2.31}$$

$$FC_i - \sum_t w_t \chi_{i,t} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{2.32}$$

Respectively, these conditions define:

- The profitability of technology i in sub-period t;
- Sub-period generation limits in terms of sub-period capacity;
- The relationship between the overall capacity selection and sub-period capacity; and
- The equilibrium investment condition in terms of sub-period profitability.

The imputed value of technology i in season t, $\chi_{i,t}$, is defined in (2.29) as the sum of profits weighted according to the respective utilisation ranges they are applicable to. A substitution from (2.26) into (2.25) gives (2.29), in which the sub-period weighted sum of $\chi_{i,t}$ defines the profit available to an

increment of capacity of technology i, which is then compared to the fixed costs of installation. $CAP_{i,t}$ is a free variable, however given $GEN_{i,k,t} \geq 0$, (2.27) implies $CAP_{i,t} \geq 0$. The relationship between $CAP_{i,t}$ and $CAP_i$ implies that $\chi_{i,t} > 0$ if and only if $CAP_{i,t} = CAP_i$. Alternatively, we can state that $\chi_{i,t} = 0$ if and only if technology i is the peaking technology in sub-period t. In economic terms, we are merely confirming that, under the assumption of marginal cost pricing, capacity is profitable in a sub-period only when the technology is infra marginal, and not profitable when the technology is marginal.

For a risk neutral investor, (2.29) defines the equilibrium relationship between fixed costs, and the imputed technology values, $\chi_{i,t}$, as well as the nature of equilibration required where the complementarity constraints are not satisfied:

- Where $CAP_i = 0$ and $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} \geq 0$ we have attained an equilibrium in which the weighted valuation of the technology across all seasons is (weakly) less than the fixed cost of installing capacity, and therefore fails to justify investment in that technology. However, if $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} < 0$, then the weighted valuation of the technology across all seasons is greater than the fixed cost of installing capacity, which incentivises further investment in technology i until $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} = 0$.
- In the case where $CAP_i > 0$ and $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} > 0$, then the weighted valuation of technology i is less than the fixed cost of installing capacity and the installed capacity must be reduced. As capacity is reduced the profitability of the technology increases, until eventually either $CAP_i = 0$ with $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} > 0$, or until $\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} = 0$ with $CAP_i > 0$. Each of these outcomes represents a legitimate equilibrium.

It is important to emphasise that the aggregate nature of the equilibrium investment decision permits each technology, whether built or not, to exhibit different levels of profitability in different seasons. In particular, (2.29) does not imply that for technologies not built $\chi_{i,t} = 0 \ \forall i > 0, t$ or $\chi_{i,t} < \mathrm{FC_i} \ \forall i > 0, t$. There may be seasons in which a particular technology i, would be infra-marginal and operating with sufficient profitability so that $\chi_{i,t} > \mathrm{FC_i}$, although not by enough, or often enough, to justify investment. Similarly, (2.29) does not imply that $\chi_{i,t} > \mathrm{FC_i} \ \forall i > 0, t$ for installed technologies, as if that is true for any season, it must be the case that $\chi_{i,t} < \mathrm{FC_i}$ in at least one season or the technology will earn supernormal profits, and further investment will be incentivised.

Were capacity flexible, the following form of investment constraint would be appropriate:

$$\mathrm{FC_i} - \chi_{i,t} = \mathrm{FC_i} - \sum_{k<K} \varphi^{+}_{i,k+1,t} \left( \mathrm{u_{k+1,t}} - \mathrm{u_{k,t}} \right) \geq 0 \quad \perp \quad CAP_{i,t} \geq 0 \qquad \forall i > 0, t \tag{2.33}$$

We can measure the value of complete flexibility, or the cost of inflexibility, by assessing the extent to which this constraint is violated in individual sub-periods. Algebraically this is:

$$\sum_{t} w_t \left| FC_i - \sum_{k<K} \varphi^{+}_{i,k+1,t} \left( u_{k+1,t} - u_{k,t} \right) \right| \qquad \forall i > 0 \qquad (2.34)$$

The difference between the weighted average of returns available to each technology in each sub-period and the average that is required for investment to occur is calculated in (2.31), and is the extent to which opportunities are missed by the need for a compromise capacity choice.

One implication of the inclusion of sub-periods is that, aside from the case where each sub-period differs only in respect to load conditions, it is no longer possible to summarise the whole period in a single screening curve diagram. This has important implications throughout the rest of the thesis, as it implies that the utilisation levels corresponding to optimal trade-offs are not consistent across sub-periods and that the utilisation of each technology cannot be averaged or assumed to have a single characteristic, such as profitability.

### 2.3.7 Solution Ambiguity

In Section 2.3.5, we presented a single period complementarity formulation and we noted that, barring coincidence, the solution would be unique. Coincidence in this context could refer to duplication of cost structures but, more relevantly, it refers to the coincidence of total cost between two technologies at a particular utilisation level. As that model is specifically designed to operate in conjunction with optimised utilisation levels corresponding to optimal trade-offs, this is no longer a coincidence, and instead is a feature of the model.

Therefore, by construction, there is ambiguity in the single period model as, at utilisation levels corresponding to optimal trade-offs, capacity choices can be varied within a certain range without affecting the optimal PDC. Where a pre-determined or optimised utilisation level corresponds to the optimal trade-off between two or more technologies then, by definition, each technology can operate at that utilisation level with the same total cost, leaving the choice of how much capacity of each should be built undetermined by the model. However, the ambiguity in capacity choices is not unlimited as it is constrained by the PDC. Capacity can only be varied to the extent that it does not result in changes in the PDC in any sub-period. But to the extent it can be varied, incremental load can be served by either technology and, according to (2.10), whichever technology or combination of the two technologies serves that load, the market price will be unaffected leaving the PDC and, by extension, the installed capacity of other technologies unaffected. This ambiguity does not register in the marginal benefit function as this is calculated on the basis that other capacity is constant, without consideration of pairwise adjustments.

Solution ambiguity arises from a practical difference between the model proposed and the logic of screening curve analysis. The very point of screening curve analysis is to define those plant roles, or utilisation levels, that correspond to where two technologies can produce at equal total cost. The definition of marginal operating ranges, and the PDC, is straight forward in the screening curve framework. While we address those same individual utilisation levels, we must explicitly state the means by which our model will interpolate those prices in the optimal investment condition. That

condition, and screening curve analysis, make clear that only one solution is consistent. The condition is valid only on the basis that there is a single marginal technology over an operating range. While interior solutions involving some capacity of each technology are not suitable, they could satisfy equilibrium conditions when viewed at a single point, although they would be inconsistent with market clearing outcomes between those points. The screening curve diagram, and our own intuition, make clear that the technology that should supply the incremental load at an optimal trade-off is the lower marginal cost technology, for it is this technology that is more efficient and operates at lower total cost at all utilisation levels, $u > u_k$ .

One approach to resolving the issue is to approximate the true solution arbitrarily closely by perturbing the utilisation level $u_k$ to $u_{k'} < u_k$ . At $u_{k'}$ , the most efficient technology to serve load is the lower fixed cost, higher variable cost technology. The optimal installed capacity of that technology will be, within a factor of the perturbation term multiplied by the slope of the LDC between $u_{k'}$ and $u_k$ , be equal to the optimal installed capacity that would be prescribed by the screening curve analysis. A more comprehensive solution that builds on the screening curve approach to force the outcome of the market clearance and investment complementarity conditions to comply with the logic of screening curve analysis, and generate a unique solution is presented in Appendix 7.2. The complementarity conditions presented therein rely on formulations yet to be presented.

One reason we have not concern ourselves greatly with a detailed resolution of this issue is that we have developed the overall framework to include sub-periods, among other divisions of the LDC. In a multiple period model, the circumstances are somewhat different. Whereas in a single period model the optimal trade-offs are defined by technological cost comparisons, that is only loosely true in multiple period formulations as while individual technologies are included in the equilibrium plant mix for broadly the same reasons, within each individual period, the PDC is actually determined by the capacity of each technology relative to the LDC, and the merit order that applies.

As we shall show in Section 2.4.2, the optimal trade-offs in each sub-period are defined in order to correctly represent the PDC. The adjustment of capacity levels, even between two technologies that are adjacent in the merit order, results in adjustments to the sub-period PDC by adjusting the utilisation levels that define it. So while prices may not adjust, the weighting associated with those prices will adjust and, in general, those adjustments will not be symmetric across sub-periods. Accordingly, there is little scope for ambiguity in capacity levels once the model evolves from being a single period model to include multiple periods.

### 2.3.8 Summary

Having established the complementarity formulation of market clearance and investment, we focus on the determination and usage of the relevant endogenous utilisation levels. The introduction of these utilisation levels is complicated by the fact that not only are they endogenous, but so are their rankings, both relative to one another, and relative to the fixed utilisation levels that we continue to include in order to define the LDC. Accordingly, we require the use of five additional sets of complementarity constraints to drive the problem representation to its optimal form:

- One set to define optimal trade-offs, as discussed in Section 2.4;
- One set to select the critical optimal trade-offs, as discussed in 2.5;
- One set to maintain the ordering and integrate endogenous utilisation levels with the fixed utilisation levels that describe the LDC, as discussed in section 2.6;
- One set to determine the load levels that correspond to those utilisation levels as discussed in section 2.6.2; and
- (Optionally) one set to resolve solution ambiguity (see Appendix 7.2)

## *2.4 Defining Optimal Trade-Offs*

### 2.4.1 Single Period Trade-Offs

In the simplest case of a single period model with deterministic cost structures, we could pre-process the data and determine optimal trade-offs before beginning other analysis. In such a simple situation, the screening curve approach is entirely tractable and no optimisation or more detailed formulation is required. In that situation, the following algorithm demonstrates one approach to defining the minimum set of optimal utilisation levels necessary for representing the system:

1. Sort technologies from highest marginal cost to lowest so that i=0, corresponds to the notional shortage technology.
2. Set i=0.
3. If $i \geq I$ go to 6, or else find

$$\underset{j>i}{Arg\ Min} = \left\{ j \middle| u_{i,j} = Min\left( \frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} \right) \right\} \tag{2.35}$$

4. If $u_{i,j} \leq 1$ record $u_{i,j}$ as a critical utilisation level and determine the corresponding load level, $L_{i,j}$ by interpolation. If $u_{i,j} > 1$ go to 6.
5. Set $i = j$, and return to step 3.
6. Stop.

This approach supposes a priori knowledge of the merit order to develop the minimum set of optimal trade-offs and therefore lacks the ability to adjust to variable merit orders, or even variable cost structures when the merit order remains unaffected. This limitation alone, rules a preparatory algorithm out of contention for analysis involving multiple sub-periods, although such an algorithm may provide a useful seed solution.

### 2.4.2 Multiple Period Trade-Offs

Even before variable, or endogenous, cost structures are considered, the introduction of sub-periods makes imputed sub-period valuations of capacity endogenous, even when the actual cost structure is not. This creates a requirement to calculate those utilisation levels that define the optimal trade-offs between technologies in each sub-period. To that end, we have identified two broad strategies:

- Market performance measures such as generation and pricing
- Pairwise technological comparison using imputed valuations

The first approach seeks to identify the critical utilisation levels for each technology using endogenously determined system performance measures. The marginal operating range for each technology is determined using system performance measures, such as profitability and generation to define when a technology enters production, or reaches full capacity. For example, the utilisation level at which each technology enters the dispatch can be determined by identifying those utilisation levels at which generation is zero and the technology is marginal, as determined by its marginal cost being equal to the market price.

The second approach identifies a superset of utilisation levels from which the critical members can be selected. By considering all pairwise trade-offs between technologies, as described in Section 2.4.4, we create a superset of utilisation levels of which the system-wide critical utilisation levels will be a subset. This approach requires development of more utilisation levels than are strictly necessary but is flexible and aligns with the intuition of the underlying economics of the problem. It could be implemented without any further refinement, although in a realistic case a significant number of the utilisation levels generated would essentially be redundant data points, for which a full market clearance would have to be calculated. The computational efficiency of the process could be further improved by ruling out options that are impossible, although the modeller should be careful not to eliminate unlikely but potentially legitimate interactions, as is the case where exogenous utilisation levels are chosen. Finally, we note that while the pairwise trade-offs are a simple function of the problem data in this simple incarnation, in more general circumstances these trade-offs will also depend on endogenous variables that are determined simultaneously. In that sense this approach shares, with the first approach, some dependence on equilibrium value of endogenous variables.

While a direct definition of the optimal trade-off between technologies i and j in each sub-period is available, the very nature of the optimal trade-off between technologies that is described by screening curve analysis is different when there are multiple operating environments under consideration. There is no longer an explicit overall utilisation level based on simple relative cost structures that corresponds to technology usage in each sub-period. The overall trade-off is implicit and is governed via the investment constraint, which aggregates individual sub-period performances. Nevertheless, we remain interested in characterising optimal trade-offs, and the most relevant optimal trade-offs are those that define the economic dispatch and correspond to steps in the price duration curve in each operating environment. Determining these utilisation levels amounts to reversing the previous logic. Rather than optimal trade-offs defining capacity, capacity defines the optimal trade-offs and PDC in each sub-period, from which investors assesses investment opportunities, using some form of weighted average of the imputed capacity valuations of each technology in each sub-period, or scenario.

For example, Figure 19 illustrates an investment decision that is based on technology performance in two sub-periods, each representing a season. The optimal capacity for each season differs, as the variable cost in each season differs. Accordingly, the desired level of capacity and the value of capacity also vary. The value of the capacity in the season represented in the left-hand pane is

lower than that in the right-hand pane, leading investors to a compromise capacity level for which the weighted average of returns equate to fixed costs.

Except for a notional shortage technology, which is entirely flexible from a capacity standpoint, all real technologies will be built to a compromise capacity level when viewed from the perspective of a single scenario or sub-period. We assume for illustrative purposes we can focus on a single technology, and discuss the adjustment of this technology without veering into the feedback issues that exist although we acknowledge that this simplification is equivalent to the contradictory assumption that the capacity of other technologies is flexible between sub-periods.

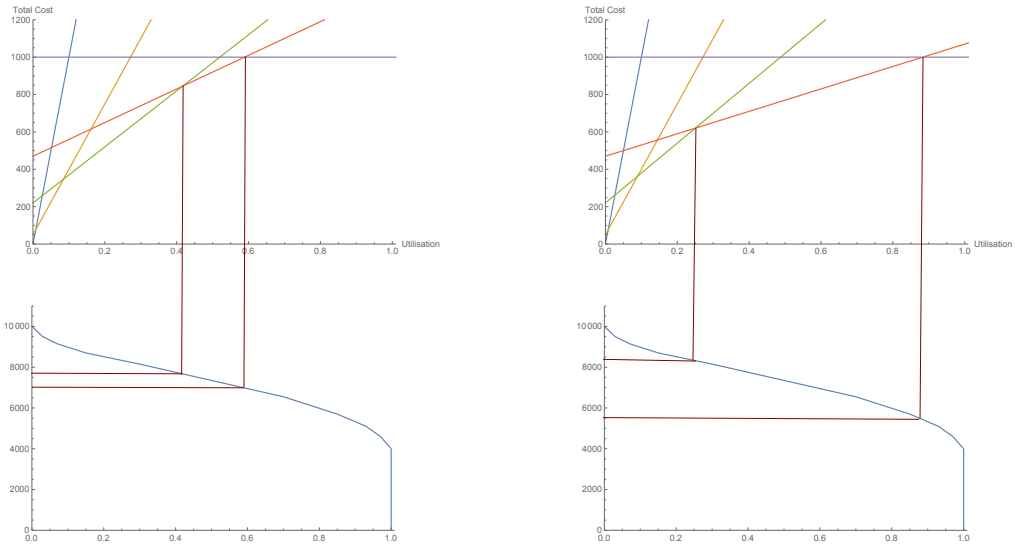



**Figure 19: Investment Compromise**

We now consider the capacity determination of a single technology i, whose marginal operating range lies between endogenous utilisation levels $u_n^e$ and $u_{n+1}^e$. While the presence of feedback loops that associate plant profitability to the capacity selection of other technologies result in a significantly larger set of simultaneous equations, we focus on the adjustment of a single technology. As the LDC is linear, with slope $\left(\mathrm{L_{k+1}}-\mathrm{L_k}\right)/\left(\mathrm{u_{k+1}}-\mathrm{u_k}\right)$, $\chi_{i,t}$ may be interpreted as providing the solution to the following simultaneous equations.

$$CAP_i = \left(\frac{\mathrm{L_{k+1}}-\mathrm{L_k}}{\mathrm{u_{k+1}}-\mathrm{u_k}}\right)\left(u_{n+1}^e - u_n^e\right) \qquad \forall i>0 \qquad (2.36)$$

$$u_{n+1}^e = \frac{\chi_{\mathrm{i+}} - \chi_{i,t}}{\mathrm{MC_i} - \mathrm{MC_{i+}}} \qquad \forall i>0,t \qquad (2.37)$$

$$u_n^e = \frac{\chi_{i,t} - \chi_{\mathrm{i-}}}{\mathrm{MC_{i-}} - \mathrm{MC_i}} \qquad \forall i>0,t \qquad (2.38)$$

Here i+, and i- refer to the neighbouring technologies as shown, which are not necessarily the technologies i+1, or i-1, from the original problem. Equation (2.36) defines the relationship between capacity and the marginal utilisation range, in terms of the (negative) slope of the LDC. As $\chi_{i,t}$ increases, $u^e_{n+1}$, the utilisation level corresponding to the load level at which technology i begins to generate falls, and $u^e_n$, the utilisation rate at which technology i is utilised at full capacity increases, thereby narrowing the width of the marginal utilisation range, $u^e_{n+1} - u^e_n$, until it is consistent, given the slope of the LDC across that utilisation range, with the capacity built. At this point the value of $\chi_{i,t}$ is the imputed value of capacity for technology i, in sub-period t, and $\chi_{i+}$ and $\chi_{i-}$ represent imputed capacity valuations for neighbouring technologies. For general, as opposed to linear, LDC forms, the intuition remains the same, although the specification of (2.36) requires modification according to the applicable functional form. The solution of this, and the other corresponding equations, generates a set of sub-period trade-offs, which define the breakpoints in the sub-period PDC.

### 2.4.3 Fixed Merit Order

To avoid potential computational issues arising from the endogenous nature of those trade-off conditions we prefer to define the optimal trade-off implicitly using $DEV_{i,j,t}$, which we define as the difference between the cost of using technology i as opposed to technology j at a given utilisation level in a given sub-period. If we consider the simultaneous adjustment of all technologies, then (2.37) and (2.38) guide us towards the following definition of $DEV_{i,j,t}$:

$$DEV_{i,j,t} = \chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^e_{i,j,t} \qquad \forall i \in I, j \in I, t \qquad (2.39)$$

Where:

$$\chi_{i,t} = \sum_{k<K} \varphi^{+}_{i,k+1,t} \left( \mathrm{u_{k+1,t}} - \mathrm{u_{k,t}} \right) \qquad \forall i,t \qquad (2.40)$$

$DEV_{i,j,t}$ is also an indicator of the degree to which the utilisation level $u^e_{i,j,t}$ is inconsistent with the utilisation level corresponding to the optimal trade-off between technologies i and j, where j>i so that $\mathrm{MC_{j,t}} < \mathrm{MC_{i,t}}$.

Where the merit-order is consistent across all sub-periods, technologies can be ordered from highest marginal cost to lowest before commencing analysis. In this case, the sign of the deviation also indicates whether the utilisation level is too high or too low, and informs the solution process accordingly.

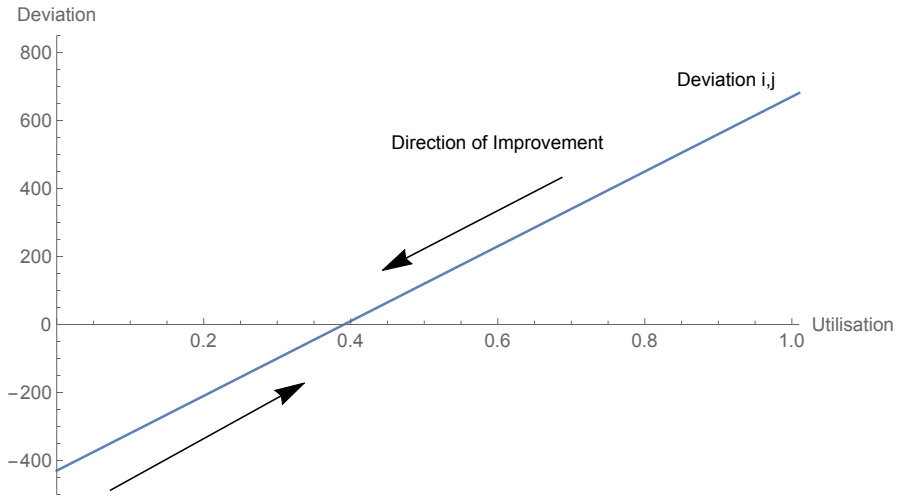



**Figure 20: Complementarity Constraints for Optimal Trade-Offs**

As shown in Figure 20, when $u^{e}_{i,j,t}$ is too high, $DEV_{i,j,t} > 0$. Conversely, when $u^{e}_{i,j,t}$ is too low, $DEV_{i,j,t} < 0$. When $0 \le u^{e}_{i,j,t} \le 1$, then $u^{e}_{i,j,t}$ represents the optimal trade-off between technology i and technology j, when $DEV_{i,j,t} = 0$. Where technological trade-offs occur outside the permissible utilisation range, $0 \le u^{e}_{i,j,t} \le 1$, the deviation cannot be driven to zero. The following complementarity conditions reflect that logic and force endogenous utilisation levels to their appropriate values where possible or, where not possible, as close as the utilisation range permits using the variable $\eta_{i,j,t}$ as a measure of the deviation between technology i and j at the utilisation level bound in sub-period t.

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \ge 0 \quad \perp \quad u^{e}_{i,j,t} \ge 0 \qquad \forall i,j > i,t \tag{2.41}$$

$$1 - u^{e}_{i,j,t} \ge 0 \quad \perp \quad \eta_{i,j,t} \ge 0 \qquad \forall i,j > i,t \tag{2.42}$$

Figure 20 also shows the direction of feasibility from the perspective of a solution algorithm such as an interior point method. Where $DEV_{i,j,t} > 0$ with $u^{e}_{i,j,t} > 0$ then infeasibility, as defined by the product of the complementary terms in (2.42), is reduced as $u^{e}_{i,j,t}$ falls until either term reaches zero, thereby satisfying the complementarity condition. Where $DEV_{i,j,t} < 0$ with $u^{e}_{i,j,t} > 0$ then, from (2.42), infeasibility is reduced as $u^{e}_{i,j,t}$ increases until either $DEV_{i,j,t} = 0$ or $u^{e}_{i,j,t} = 1$, at which point $\eta_{i,j,t} > 0$ from (2.42) satisfies the complementarity constraint (2.41).

### 2.4.4 Variable Merit Order

Unfortunately, it is not generally possible to rely on the stability of the merit order, as the very phenomena that are often central to the analysis, such as inflows in a hydro setting or the incidence of carbon taxation, will typically also drive variation in the merit order. Ideally, we would only determine the utilisation levels corresponding to the "significant" technological trade-offs but, unless some

simplifying assumptions can be made, we must consider a wider range of utilisation levels. Therefore, although we can rely on the position of the notional shortage technology, when the merit order is variable we must consider all pairwise optimal trade-offs to guarantee all critical points are captured, as the set of relevant technologies is not necessarily consistent between scenarios or sub-periods. In fact, we cannot even rely on the consistency of the pairwise ordering of that set, as fuel cost variations and energy limits, for example, could swap the merit order position of two technologies.

The requirement to consider additional utilisation levels presents algorithmic difficulties, in addition to the extra computational burden imposed. The definitions offered in (2.41) and (2.42) no longer necessarily characterise the optimal trade-offs required when the merit order is not fixed. From Figure 21, consider technology A, which has a higher marginal cost than technology B, and $DEV_{A,B} > 0$ at $u_k$. As in Figure 20, satisfaction of the complementarity condition requires an adjustment to the utilisation level $u_{A,B}$ where $DEV_{A,B} = 0$. However, when the relative cost rankings are reversed $DEV_{A,B} < 0$, the direction of decreasing infeasibility is reversed and we achieve feasibility by adjustment so that $u_{A,B} = 1$, and complementarity is restored with $\eta_{A,B} > 0$. The cost structure of two technologies relative to one another determines the direction of decreasing infeasibility and when it switches, so does the utilisation level that satisfies the complementarity conditions (2.41) and (2.42). Rectifying the situation requires consideration of all pairwise comparisons in both directions, as defined below:

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (2.43)$$

$$1 - u^{e}_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (2.44)$$

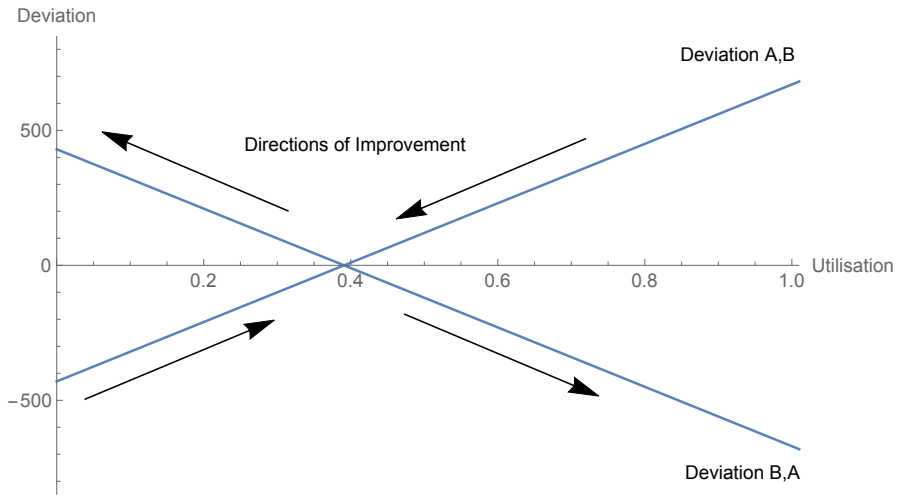



**Figure 21: Convergence Directions with Variable Cost Structures**

For each pairwise comparison between technologies i and j, the utilisation level that satisfies the complementarity constraints (2.43) and (2.44) will be satisfied by a (weakly) interior solution when the

trade-off is defined in one direction, (i,j), while in the other direction, (j,i), adjustment to a redundant utilisation level at zero or one will be the outcome. A priori, it is often not possible to determine which direction corresponds to the actual orientation of an optimal trade-off, so this procedure requires the identification of twice as many utilisation levels as are required when the merit order is fixed.

Reduction in the number of pairwise trade-offs to be considered could be achieved by a priori identification of those trade-offs that can be practically ignored on the basis of system knowledge. Among the critical utilisation levels that could be eliminated would be those corresponding to trade-offs between technologies that could only ever be in one direction, as well as trade-offs that could never eventuate in either direction because of the presence of intermediate technologies under all (modelled) circumstances. As the number of pairwise comparisons increases, the temptation to eliminate certain possibilities also increases but counter-balancing the temptation of a reduced computational burden is the risk that we may pre-suppose the solution, based on a potentially prejudiced view of the typical roles and capabilities of various technologies. In the following section, we consider alternative approaches designed to remove the judgement of the modeller from that process. The approaches discussed represent an intermediate step in that they select critical utilisation levels from a complete set of optimal trade-offs, which is less desirable that ruling out potential trade-offs from the outset, but still beneficial as it reduces the number of instances market clearances need to be calculated.

## *2.5 Selecting Critical Utilisation Levels*

Given any two technologies, one or both technologies may be dominated and not feature in the optimal plant mix at all, and even when both technologies do feature, they may not be adjacent to each other in the merit order. Thus, the set of utilisation levels corresponding to all pairwise trade-offs is a superset of the critical utilisation levels we require. When the number of technologies is large, it may result in the consideration of a significant number of redundant utilisation levels. Optionally, we may seek to determine only those members of the superset that are the critical utilisation levels that describe the screening curve lower envelope.

### 2.5.1 Pruning Utilisation Levels

In Section 2.5, we described how a simple algorithm would choose important utilisation levels by effectively navigating the lower envelope of the screening curve. We now explore how we could integrate this functionality into a complementarity formulation. One approach is to prune the set of optimal trade-offs by selecting only those utilisation levels at which each technology is superseded by the next, more efficient, technology. This selection process could reflect trade-offs with either the immediately higher or lower utilisation technology, and would work if that choice is consistently applied.

Instead of defining the market clearing complementarity conditions relative to the set of all $u^{e}_{i,j,t}$, we define an intermediate variable $u^{e}_{i,t}$, corresponding to the level at which each technology i is superseded in sub-period t, using the following set of complementarity constraints:

$$\sum_j \psi_{i,j,t} - 1 \geq 0 \quad \perp \quad u^e_{i,t} \geq 0 \qquad \forall i,t \qquad (2.45)$$

$$u^e_{i,j,t} - u^e_{i,t} \geq 0 \quad \perp \quad \psi_{i,j,t} \geq 0 \qquad \forall i,j>i,t \qquad (2.46)$$

Here $u^e_{i,t} \leq u^e_{i,j,t} \; \forall i,j,t$. (2.45) requires $\psi_{i,j,t} > 0$, at least once for each technology in each sub-period, so by complementarity we require $u^e_{i,t} = u^e_{i,j,t}$ for that combination, which by definition must be the minimum $u^e_{i,j} \; \forall i,j$, and need not be unique in the case where multiple technologies trade-off at the same point with technology i. By construction, utilisation levels are restricted to the unit interval.

The result is that the number of constraint instances in the main body of the formulation is reduced so that it contains only a single utilisation level corresponding to each technology i. If we treat this simple model as a single scenario case, with I+1 technologies, then for each scenario there are $\mathrm{I}^2 + \mathrm{I}$ pairwise comparisons to be made, which after processing with (2.45) and (2.46) leads to I+1 endogenous utilisation levels being calculated and used to enhance the LDC representation, which is $\mathrm{I}^2 - 1$ fewer than the full set would represent. We can gauge whether or not the application of complementarity conditions for pre-processing constraints such as (2.45) and (2.46) delivers a net computational benefit. To incorporate each of these complementarity conditions adds $\left(\mathrm{I}^2 + \mathrm{I}\right)/2$ sets of conditions for a total of $\mathrm{I}^2 + \mathrm{I}$ extra conditions. The benefit is measured in the reduction of instances of other complementarity conditions. For each condition indexed by utilisation levels, there will be $\mathrm{I}^2 - 1$ multiplied by the dimensionality of the other indices involved fewer instances of those conditions. For example, a market clearing condition such as (2.21) would have $\left(\mathrm{I}+1\right)\left(\mathrm{I}^2 - 1\right)$ fewer instances. For all but a trivial system with a single technology, the additional pre-processing conditions represent fewer additional conditions than the reduction in instances of a single condition in the larger model. As there are many conditions linked to utilisation levels, and sometimes with far greater dimensionality, it seems clear that the computational benefits of pre-processing are significant.

This approach is broadly equivalent to the pre-processing algorithm in Section 2.4.1, in that fewer utilisation levels, $u^e_i$, are to be considered by the main optimisation or other complementarity conditions. The number of constraint instances, or critical utilisation levels, for the investment and market clearance problem is equal to I+1, the number of technologies, which is weakly greater than the number of critical utilisation levels generated by the algorithm, as the algorithm avoids specifying redundant utilisation levels by resetting i each step.

### 2.5.2 Defining the Screening Curve Lower Envelope

It is possible to refine the above approach further and more precisely determine endogenous utilisation levels. The approach that follows is unlikely to be computationally beneficial, although in cases where market-clearing conditions are sufficiently onerous the addition of additional complementarity constraints could be preferable to additional instances of market clearing constraints. This is

particularly so when the problem is solved in a decomposed form, or when technological cost functions are not affine.

Adopting the approach of the algorithm described in Section 2.4.1, this formulation determines the lower envelope of the screening curve iteratively and defines the technology optimally used for each segment. Starting with the notional shortage technology, we progress to find the next minimum intersection, and the technology that it represents. We continue to explore the lower envelope in the direction of increasing utilisation to find the next intersection. Each search is an optimisation of its own, in which the previous solution is assumed fixed.

Accordingly, we propose a series of nested optimisation problems, each choosing the next utilisation level according to the next available trade-off in an endogenously constructed merit order, thereby defining the marginal operating range applicable to each single technology in each sub-period. We identify N+1 endogenous utilisation levels, indexed by n=0…N, where N can be problem, or even sub-period, specific and can be tuned by the user to correspond to the number of segments required to define the lower envelope. N should be large enough to fully describe that envelope without being so large that an excessive number of redundant points corresponding to full utilisation require evaluation. With constant marginal costs, N need be no larger than I+1, but could be smaller than the maximum number of intersections if redundant technologies exist in certain sub-periods.

Complementarity conditions are required to incentivise these movements, as well as deal with potential degeneracy in the form of technological cycling, and to prevent non-integer, or mixed, solutions that would lead to incorrect outcomes. These complementarity conditions use the $u^{e}_{i,j,t}$ values determined in the optimal trade-off section, to form a minimal set of critical utilisation levels on which the market clearance and optimisation of investment may occur.

We also introduce $z_{i,n,t} \geq 0 \ \forall i,n,t$ which assumes the value $z_{i,n,t} = 1$ when technology i is marginal in lower envelope segment n in sub-period t, as defined by the range $u^{e}_{n,t}$ to $u^{e}_{n+1,t}$, and $z_{i,n,t} = 0$ elsewhere. To seed the recursive procedure, we introduce the following initial conditions:

- $u^{e}_{0,t} = 0 \ \forall t$, to signify the first critical utilisation level in each sub-period, and;
- $z_{0,0,t} = 1 \ \forall t$ and $z_{0,i,t} = 0 \ \forall i \neq 0,t$, signifying the first segment of the lower envelope in each sub-period corresponds to the notional shortage technology.

We define n=0…N optimisation problems, with the nth optimisation using the incoming technology choice as an input. In each optimisation, we seek to minimise $u^{e}_{n,t}$. From the set of utilisation levels consistent with the incoming technological choice, $z_{j,n,t}$ is chosen to select the next marginal technology j and minimise $u^{e}_{n,t}$. The expression $\sum_{i} \mathrm{z_{i,n-1,t}u^{e}_{i,j,t}}$ identifies a vector of j potential trade-offs that can be chosen given the vector $\mathrm{z_{n-1,t}}$, which defines the incoming technological choice in sub-period t. We also note (2.48) and $z_{j,n,t} \geq 0$ collectively define $u^{e}_{n,t}$ as a convex combination of

technological trade-offs that is restricted to the range $0 \le u^e_{n,t} \le 1$ by $0 \le u^e_{i,j,t} \le 1$. To select the minimum positive value $u^e_{n,t}$, with $0 \le u^e_{n,t} \le 1$, $z_{j,n,t}$ is naturally integer valued whenever the utilisation levels corresponding to technological trade-offs are distinct.

$$u^e_{n,t} = \sum_j z_{j,n,t} \sum_i \mathrm{z_{i,n\text{-}1,t}\, u^e_{i,j,t}} \qquad :\psi^0_{n,t} \qquad \forall n > 0, t \qquad (2.47)$$

$$\sum_j z_{j,n,t} \ge 1 \qquad :\psi^1_{n,t} \qquad \forall n > 0, t \qquad (2.48)$$

As defined above, there is no restriction preventing the selection of $u^e_{i,j,t} \le u^e_{n,t}$. Even if we could force progression by implementing a strict inequality constraint it would not resolve misspecification of the lower envelope when the total cost of two technologies coincide at $u^e_{n-1,t}$ and, as a result of an arbitrary algorithmic decision, the lower marginal cost technology is not chosen, and the wrong trade-offs are sought from that point onwards. As can be seen from that example, the requirement is not to advance the utilisation level per se, but also to advance the discovery of the merit order in a systematic fashion that avoids the pitfalls of degeneracy.

$$\sum_j \mathrm{z_{j,n\text{-}1,t}\, MC_{j,t}} - \sum_j z_{j,n,t} \mathrm{MC_{j,t}} \ge 0 \qquad :\psi^2_{n,t} \qquad \forall n > 0, t \qquad (2.49)$$

Provided $z_{j,n,t}$ is integer valued, (2.49) ensures that the selection of the next technology is consistent with the merit order position of that technology and, by discerning on the basis of marginal cost, it prevents indefinite cycling between two or more technologies that produce at equal total cost at $u^e_{n,t}$. However, as (2.49) is an inequality, it permits the repeated choice of the same technology, or more generally any technology with the same marginal cost. We must not permit that repetition of choice so we require the inner product of two adjacent technology selection vectors must also be zero.

$$-\sum_j \mathrm{z_{j,n\text{-}1,t}}\, z_{j,n,t} \ge 0 \qquad :\psi^3_{n,t} \qquad \forall n > 0, t \qquad (2.50)$$

The exception to this rule is when the utilisation level $u^e_{n,t} = 1$. At such a time, we require the algorithm to cycle repeatedly for two reasons. Firstly, we are not interested in generating utilisation levels in excess of unity, and secondly, it may be the case that no technological options with lower marginal costs remain. Given (2.48) requires a selection be made so that $z_{j,n,t} > 0$ for some technology j, then under the circumstances the only solution to (2.49) is to select $z_{j,n,t} = z_{j,n-1,t}$, which is explicitly prohibited by (2.50). To allow this constraint to be broken when $u^e_{n,t} = 1$ requires the following modification to (2.50), along with the addition of a complementarity condition:

$$-\sum_j \mathrm{z_{j,n\text{-}1,t}}\, z_{j,n,t} + \psi^4_{n,t} \ge 0 \qquad :\psi^3_{n,t} \qquad \forall n > 0, t \qquad (2.51)$$

$$1 - u^e_{n,t} \ge 0 \quad \perp \quad \psi^4_{n,t} \ge 0 \qquad \forall n > 0, t \qquad (2.52)$$

The last remaining issue is when two technologies define the next intersection, at the same utilisation level. The condition (2.49) appears to maintain the consistency of the merit order but it is applied retrospectively, not prospectively. Therefore, when two technologies represent equal total cost at the next selected critical utilisation level there is no guarantee that the ordering based on marginal cost will be selected. Mathematically we have multiple solutions to the optimisation, which are unique only up to the convex combination of those two technologies. Where $z_{j,n,t}$ is not integer valued, such as when more than one technology produces with equal total cost at $u^e_{n,t}$, an arbitrary convex combination of two technologies will result. The search for $u^e_{n+1,t}$ will then be based on a non-existent technological composite of two technologies with a cost structure that is a convex combination of the cost structures of two real technologies. This will lead to an incorrect evaluation of $u^e_{n+1,t}$ and depending on the actual combination chosen, possibly also an incorrect identification of the marginal technology between $u^e_{n,t}$ and $u^e_{n+1,t}$. The anti-cycling condition (2.49) does not prevent this eventuality and instead, in combination with (2.50), compounds potential problems by ruling all technologies involved in that convex combination out of further consideration. The requirement to select only a single technology is enforced by the following complementarity condition:

$$\sum_j {z_{j,n,t}}^2 - 1 \geq 0 \qquad :\psi^5_{n,t} \qquad \forall n > 0, t \tag{2.53}$$

As $0 \leq z_{j,n,t} \leq 1$, the maximum value of the LHS of (2.53) is 0 when $z_{j,n,t} = 1$ for a single j. Where more than one technology produce at equal total cost at $u^e_{n,t}$, the selection of either could satisfy this constraint, although it matters little to the formulation. If the higher marginal cost technology is selected, it will then be eliminated by (2.50) from further consideration and the next $u^e_{n,t}$ will take the same value, although the selected technology will now be the lower marginal cost technology, which is the appropriate choice from which to base a search for $u^e_{n+1,t}$. If the lower marginal cost technology is chosen first, then the search for a successive $u^e_{n,t}$ proceeds as normal. As the formulation must already generate several redundant utilisation levels, the only implication for technology choice is where those redundant utilisation levels will appear. The requirement that the successive marginal costs of each technology selected must be monotone non-increasing as expressed in (2.49), in combination with a restriction on selecting the same technology consecutively, (2.50), provides strong protection against the potential issue of cycling.

Finally, we must ensure that we only select those technologies that are built when forming the screening curve envelope. The definition of optimal trade-offs relies on the imputed capacity value of a technology in a sub-period, which suggests individual technologies that are not built may interfere with the process, as it is only on average that they are not required. The following constraint prevents the selection of a technology that is not built:

$$\mathrm{CAP}_{\mathrm{j,t}} - z_{j,n,t} \geq 0 \qquad :\psi^6_{j,n,t} \tag{2.54}$$

Where the capacity of a candidate technology is zero, then $z_{j,n,t} = 0$. Where capacity is positive, and very small, there is also no possibility of selecting the technology as $z_{j,n,t} < 1$ which is prohibited by (2.53). This results in a slight inaccuracy, albeit a helpful one, as it eliminates from consideration an optimal trade-off corresponding to a technology that has an impractical level of capacity, and therefore would not be built. This effect could be enhanced by introducing a scaling factor into (2.54) to regulate a minimum capacity level, or more exactly, a minimum step-size in the screening curve lower envelope, that might lead to unrealistically small technologies being overlooked on account of the system definition. Finally, we note that where technologies are built but not used in a sub-period, they will be considered, although they will cycle through the process defining a section of the envelope with zero length.

Combining the above constraints and complementarity condition we have a series of N optimisation problems for each sub-period t, specified as below for a particular (n>0,t) combination:

Minimise: $u^e_{n,t}$

Subject to:

$$u^e_{n,t} - \sum_j z_{j,n,t} \sum_i \mathrm{z}_{\mathrm{i,n\text{-}1,t}} \mathrm{u}^{\mathrm{e}}_{\mathrm{i,j,t}} \geq 0 \qquad :\psi^0_{n,t} \qquad \forall n,t \qquad (2.55)$$

$$\sum_j z_{j,n,t} - 1 \geq 0 \qquad :\psi^1_{n,t} \qquad \forall n,t \qquad (2.56)$$

$$\sum_j \mathrm{z}_{\mathrm{j,n\text{-}1,t}} \mathrm{MC}_{\mathrm{j,t}} - \sum_j z_{j,n,t} \mathrm{MC}_{\mathrm{j,t}} \geq 0 \qquad :\psi^2_{n,t} \qquad \forall n,t \qquad (2.57)$$

$$-\sum_j \mathrm{z}_{\mathrm{j,n\text{-}1,t}} z_{j,n,t} + \psi^4_{n,t} \geq 0 \qquad :\psi^3_{n,t} \qquad \forall n,t \qquad (2.58)$$

$$1 - u^e_{n,t} \geq 0 \qquad :\psi^4_{n,t} \qquad \forall n,t \qquad (2.59)$$

$$\sum_j z_{j,n,t}{}^2 - 1 \geq 0 \qquad :\psi^5_{n,t} \qquad \forall n,t \qquad (2.60)$$

$$\mathrm{CAP}_{\mathrm{j,t}} - z_{j,n,t} \geq 0 \qquad :\psi^6_{j,n,t} \qquad \forall j,n,t \qquad (2.61)$$

$$z_{j,n,t} \geq 0 \qquad \forall j,n,t \qquad (2.62)$$

Combining the constraints (2.55)-(2.61) and first order conditions from all N instances of this problem, with the initial conditions specified above, we have the following representation of the lower envelope of cost curves in the form of a set of complementarity conditions. In forming the first order conditions for each individual problem, we treat the vector $z_{n-1,t}$ as a constant, having been optimally determined by the previous set of conditions. As we now define a group of equations $z_{n-1,t}$ reverts to being a variable, although the cascading nature the optimisation problems remains.

First Order Conditions:

$$1-\psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.63)$$

$$\psi^{0}_{n,t}\sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC}_{\mathrm{j,t}} + \psi^{3}_{n,t} z_{j,n\text{-}1,t} - 2\psi^{5}_{n,t} z_{j,n,t} + \psi^{6}_{j,n,t} \geq 0$$

$$\perp \quad z_{j,n,t} \geq 0 \qquad \forall j,n>0,t \qquad (2.64)$$

Constraints:

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t} \sum_{i} z_{i,n\text{-}1,t} u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.65)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.66)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC}_{\mathrm{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC}_{\mathrm{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.67)$$

$$-\sum_{j} z_{j,n\text{-}1,t} z_{j,n,t} + \psi^{4}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.68)$$

$$1-u^{e}_{n,t} \geq 0 \quad \perp \quad \psi^{4}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.69)$$

$$\sum_{j} z_{j,n,t}{}^{2} - 1 \geq 0 \quad \perp \quad \psi^{5}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (2.70)$$

$$\mathrm{CAP}_{\mathrm{j,t}} - z_{j,n,t} \geq 0 \quad \perp \quad \psi^{6}_{j,n,t} \geq 0 \qquad \forall j,n>0,t \qquad (2.71)$$

Initial Conditions:

$$u^{e}_{0,t} = 0 \qquad \forall t \qquad (2.72)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (2.73)$$

$$z_{0,i,t} = 0 \qquad \forall i>0,t \qquad (2.74)$$

By inspection the system is square, having the same number of variables as conditions. From this point forward, we continue to use $u^{e}_{n,t}$ as notation for a critical endogenously determined utilisation level, however we wish to emphasise that in what follows that utilisation level could have been developed by some other means, such as the approach discussed in Section 2.5.

## *2.6 LDC Construction*

### 2.6.1 Ordering Utilisation Levels

While the solution to the optimal plant mix problem depends on the calculation of the critical utilisation levels at which optimal trade-offs are defined, there also remains a need to incorporate

exogenous utilisation levels in the model, for the purpose of defining the LDC as input. The structure of the original optimisation problem, and the resulting complementarity conditions detailed by (2.21)-(2.24), require utilisation levels to be considered in a monotonic fashion if constraints, objectives, and first order conditions are to make any sense. Assuming we have already determined the appropriate settings for utilisation levels corresponding to the optimal technological trade-offs, we now need to determine the appropriate ordering of the combined set of utilisation levels.

Fortunately, a simple linear programming solution that provides a naturally integer solution to the problem of ordering distinct utilisation levels is available. For each sub period t, the formulation below provides a combined ordering of exogenous (k) and endogenous (n) utilisation levels, indexed by r=1….R. Using a rank based weighting scheme incentivises the optimisation to associate higher valued utilisation levels with higher rankings, thereby providing variables $x_{k,r,t}$ and $x^{e}_{n,r,t}$ to translate between the original indices and the combined ranking:

$$Minimise \; -\sum_{r}\left(\sum_{k} \mathrm{r.u}_{\mathrm{k,t}} x_{k,r,t} + \sum_{n} \mathrm{r.u}^{\mathrm{e}}_{\mathrm{n,t}} x^{e}_{n,r,t}\right) \tag{2.75}$$

Subject to:

$$1-\sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \qquad :\phi^{0}_{r} \qquad \forall r \tag{2.76}$$

$$1-\sum_{r} x_{k,r,t} \geq 0 \qquad :\phi^{f}_{k} \qquad \forall k \tag{2.77}$$

$$1-\sum_{n} x^{e}_{n,r,t} \geq 0 \qquad :\phi^{e}_{n} \qquad \forall n \tag{2.78}$$

$$x_{k,r,t}, x^{e}_{n,r,t} \geq 0 \qquad \forall k,n,r \tag{2.79}$$

Here $\mathrm{u}_{\mathrm{k,t}}$ represents a breakpoint in the piecewise linear representation of the LDC, while $u^{e}_{n,t}$ represents the critical utilisation level indexed by n in sub-period t. The objective function maximises the sum of all weighted utilisation levels, where the weightings are given by either $x_{k,r,t}$ or $x^{e}_{n,r,t}$. The highest weighting, r=R, is attached to the highest utilisation level, by selecting $x_{k,R,t}=1$ or $x^{e}_{n,R,t}=1$ for whichever $\mathrm{u}_{\mathrm{k,t}}$ or $u^{e}_{n,t}$ corresponds to the maximum utilisation level in either set. With the exception of coincidentally identical utilisation levels, constraints (2.76) to (2.79) ensure a unique transformation between the unranked and ranked utilisation levels, as each $\mathrm{u}_{\mathrm{k,t}}$ and $u^{e}_{n,t}$ is allocated a ranking in the combined set, and all rankings in the combined set correspond to a unique $\mathrm{u}_{\mathrm{k,t}}$ or $u^{e}_{n,t}$. In the case where there are coincidentally identical utilisation levels, the solution to the ranking problem above is not unique, or naturally integer. But it matters little what weightings are attributed to two identical outcomes, as they are required to sum to unity and, as is the case here, the wider problem is only concerned with convex combinations of outcomes, both of which are identical in this case. Casting these t sub-period problems as complementarity conditions we have:

First Order Conditions:

$$-\mathrm{r}.\mathrm{u}_{\mathrm{k,t}} + \phi^{0}_{r,t} + \phi^{f}_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k,r,t \qquad (2.80)$$

$$-\mathrm{r}.u^{e}_{n,t} + \phi^{0}_{r,t} + \phi^{e}_{n,t} \geq 0 \quad \perp \quad x^{e}_{n,r,t} \geq 0 \qquad \forall n,r,t \qquad (2.81)$$

Constraints:

$$1 - \sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{0}_{r,t} \geq 0 \qquad \forall r,t \qquad (2.82)$$

$$1 - \sum_{r} x_{k,r,t} \geq 0 \quad \perp \quad \phi^{f}_{k,t} \geq 0 \qquad \forall k,t \qquad (2.83)$$

$$1 - \sum_{r} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{e}_{n,t} \geq 0 \qquad \forall n,t \qquad (2.84)$$

The variables $x_{k,r,t}$ and $x^{e}_{n,r,t}$ that define the solution to this complementarity problem directly enable the transformation between unranked and ranked versions of the variables in both directions as follows:

$$u^{rank}_{r,t} = \sum_{k} x_{k,r,t}\mathrm{u}_{\mathrm{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} \qquad \forall r,t \qquad (2.85)$$

$$\mathrm{u}_{\mathrm{k,t}} = \sum_{r} x_{k,r,t} u_{r,t} \qquad \forall k,t \qquad (2.86)$$

$$u^{e}_{n,t} = \sum_{r} x^{e}_{n,r,t} u_{r,t} \qquad \forall n,t \qquad (2.87)$$

Finally, we note that the efficiency of the sorting algorithm could be improved by removing the end points $\mathrm{u}_{0,\mathrm{t}} = 0$ and $\mathrm{u}_{\mathrm{K,t}} = 1$ from this optimisation, and defining them directly in the basic formulation. This ranking scheme is also easily generalised to enable combined rankings of any number of indices should an additional set of utilisation levels be of interest for some other reason.

### 2.6.2 Defining Load at Critical Utilisation Levels

For the market clearing and investment conditions to make sense, individual load levels must also be ranked consistently according to the ranking of the corresponding utilisation level. This must occur for each sub-period. As with utilisation levels, the ranked load levels, $L_{r,t}$, corresponding with $u_{r,t}$ are analogously defined by the expression:

$$L_{r,t} = \sum_{k} x_{k,r,t}\mathrm{L}_{\mathrm{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} \qquad \forall r,t \qquad (2.88)$$

The load levels, $\mathrm{L}_{\mathrm{k,t}}$, corresponding to each fixed utilisation level, $\mathrm{u}_{\mathrm{k,t}}$, are known, and will be correctly ranked by the ranking variables $x_{k,r,t}$. In terms of the endogenous load levels, the ranking variable, $x^{e}_{n,r,t}$, can also be relied upon to correctly rank the endogenous load level, $L^{e}_{n,t}$, however $L^{e}_{n,t}$, the load corresponding to $u^{e}_{n,t}$, has not yet been defined.

Geometrically, we can define intermediate load levels such as $L^e_{n,t}$ by interpolation. Mathematically this is represented by the following expression:

$$L^e_{n,t} = L_{r-1,t} - \left(u^e_{n,t} - u_{r-1,t}\right)\frac{L_{r-1,t} - L_{r+1,t}}{u_{r+1,t} - u_{r-1,t}} \qquad \forall n,r,t \tag{2.89}$$

Unfortunately, there is no guarantee that the adjacent load levels are known, as they may also correspond to endogenous utilisation levels. Eventually though, any set of interior endogenous load levels, for example {B,C}, will be encased by two exogenous load levels, {A,D} even if, in the limit, these correspond to load at $u_{0,t}$ and $u_{K,t}$. There is only one possible consistent linear interpolation, between the two exogenous load and utilisations levels A and D. If we assume a load for C then the interpolation between A and C will uniquely define B. Accordingly, C will be uniquely defined by the interpolation between B and D, which we know from basic geometry has a single solution. The only consistent interpolation between the two fixed points is a straight line between A and D, which makes all interpolations consistent. We can represent this situation with the system of linear equations shown below:

$$L^e_B\left(u^e_C - u_A\right) - L^e_C\left(u^e_B - u_A\right) = L_A\left(u^e_C - u^e_B\right) \tag{2.90}$$

$$L^e_C\left(u_D - u^e_B\right) - L^e_B\left(u_D - u^e_C\right) = L_D\left(u^e_C - u^e_B\right) \tag{2.91}$$

Solving (2.90) for $L^e_B$ gives:

$$L^e_B = \frac{L_A\left(u^e_C - u^e_B\right) + L^e_C\left(u^e_B - u_A\right)}{\left(u^e_C - u_A\right)} \tag{2.92}$$

Substituting (2.92) into (2.91) and grouping like terms gives the following equation defining $L^e_C$:

$$L^e_C\left[\left(u_D - u^e_B\right) - \frac{\left(u^e_B - u_A\right)\left(u_D - u^e_C\right)}{\left(u^e_C - u_A\right)}\right] = L_D\left(u^e_C - u^e_B\right) + L_A\frac{\left(u^e_C - u^e_B\right)\left(u_D - u^e_C\right)}{\left(u^e_C - u_A\right)} \tag{2.93}$$

This can be simplified to the following form:

$$L^e_C\left(u_D - u_A\right)\left(u^e_C - u^e_B\right) = L_D\left(u^e_C - u^e_B\right)\left(u^e_C - u_A\right) + L_A\left(u^e_C - u^e_B\right)\left(u_D - u^e_C\right) \tag{2.94}$$

With further cancellation and substitution into (2.92) we have:

$$L^e_C = L_D\frac{\left(u^e_C - u_A\right)}{\left(u_D - u_A\right)} + L_A\frac{\left(u_D - u^e_C\right)}{\left(u_D - u_A\right)} \tag{2.95}$$

$$L^e_B = L_D\frac{\left(u^e_B - u_A\right)}{\left(u_D - u_A\right)} + L_A\frac{\left(u_D - u^e_B\right)}{\left(u_D - u_A\right)} \tag{2.96}$$

Both $L_B^e$ and $L_C^e$ and, more generally, any load levels corresponding to utilisation levels between $u_A$ and $u_D$, can be expressed as weighted averages of, and therefore linear interpolations between, the exogenous endpoints, $L_A$ and $L_D$.

While the definition in (2.89) suffices whenever $u_{r+1,t} - u_{r-1,t} \neq 0$, when the denominator is zero the solution is undefined. This will only be the case where three or more consecutive utilisation levels are identical. This might be thought unlikely, but in general there will be less than N+1 distinct utilisation levels required to determine the screening curve lower envelope implying there is likely to be a redundancy at full utilisation. For example, if E utilisation levels are required to define the lower envelope, then barring coincidental intersections in the determination of the screening curve envelope, then N+1-E utilisation levels will be set to unity, causing the denominator in (2.89) to equal zero and the interpolation method to fail whenever N+1-E>1.

This realisation also raises a broader issue of algorithm design that requires the correct choice of starting and intermediate values to avoid the denominator evaluating at zero at any point in the solution process, and not just at the final solution. With some algorithmic modifications, such as use of perturbations, these difficulties may be able to be overcome in a practical sense. To avoid this potential difficulty with certainty, we define load levels without reference to neighbouring load and utilisation levels that may be identical. The solution to this problem provides an example of a useful complementarity formulation for a particular type of constraint. Were the LDC convex, a simple optimisation such as below would suffice for each sub-period t, as incentives would ensure the ordered filling of the LDC.

$$Minimise \;\; L_{n,t}^e = \mathrm{L}_{0,\mathrm{t}} - \sum_k \frac{\mathrm{L}_{\mathrm{k\text{-}1,t}} - \mathrm{L}_{\mathrm{k,t}}}{\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}}} u_{k,n,t}^{part} \tag{2.97}$$

$$\text{Subject to:} \;\; \sum_k u_{k,n,t}^{part} = u_{n,t}^e \qquad \forall n \tag{2.98}$$

$$u_{k,n,t}^{part}, u_{n,t}^e \geq 0 \qquad \forall k, n \tag{2.99}$$

The utilisation level $u_{n,t}^e$ is allocated amongst the k load classes in quantities $u_{k,n,t}^{part}$. Each of these load classes has a prevailing linear adjustment as a function of utilisation, with the rate measured by the ratio $\left(\mathrm{L}_{\mathrm{k\text{-}1,t}} - \mathrm{L}_{\mathrm{k,t}}\right) / \left(\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}}\right)$. Convexity of the LDC implies that load classes would be filled in logical order until the total required utilisation level had been satisfied.

However, our problem amounts to assessing the value of the LDC at a particular utilisation level when the LDC is represented by a non-convex piecewise linear function. As the LDC is non-convex, incentives alone will not result in the LDC load classes will be filled in an ordered fashion, and therefore also do not guarantee that the correct corresponding load level value will be calculated. To rectify this problem, we add the following complementarity constraint, which recursively ensures that no load class can be used unless the previous load class has been entirely used:

$$\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}} - u_{k,n,t}^{part} \geq 0 \quad \perp \quad u_{k+1,n,t}^{part} \geq 0 \qquad \forall 0 < k < K, n \tag{2.100}$$

Having added this constraint, the problem above contains only one feasible point, and therefore the optimisation paradigm is redundant. Nevertheless, the objective function serves as a useful definition of the load level we seek and, together with the constraint (2.98) and the complementarity condition (2.100), $L^{e}_{n,t}$ is defined for a given endogenous $u^{e}_{n,t}$. In aggregate, the definition of endogenous load levels and ordering of the combined set of load levels requires the following set of equations and complementarity conditions:

$$L_{r,t} - \sum_{k} x_{k,r,t} \mathrm{L}_{\mathrm{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r,t \qquad (2.101)$$

$$L^{e}_{n,t} - \mathrm{L}_{0,\mathrm{t}} + \sum_{k} \frac{\mathrm{L}_{\mathrm{k\text{-}1,t}} - \mathrm{L}_{\mathrm{k,t}}}{\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n,t \qquad (2.102)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n,t \qquad (2.103)$$

$$\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall k < K, n,t \qquad (2.104)$$

## *2.7 Solution Approaches*

### 2.7.1 Complementarity Solution

As designed the formulation is a complete complementarity problem. Several substitutions are advisable to reduce the problem to its canonical form, to improve solver performance. We state the complementarity problem, including endogenous utilisation levels, necessary re-ordering conditions, and linking equations. Note that by assumption costs are fixed across all sub-periods and this is reflected in the indexation of those variables and constraints where it is appropriate.

Market Equilibrium Conditions:

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} + \varphi^{+}_{i,r,t}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (2.105)$$

$$\sum_{i} GEN_{i,r,t} - L_{r,t} = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \qquad (2.106)$$

$$CAP_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,r,t} \geq 0 \qquad \forall i > 0, r,t \qquad (2.107)$$

Optimal Capacity Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi^{+}_{i,r+1,t} \left( u_{r+1,t} - u_{r,t} \right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i,t \qquad (2.108)$$

$$CAP_{i} - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (2.109)$$

$$\mathrm{FC}_{\mathrm{i}} - \sum_{t} \mathrm{w}_{\mathrm{t}} \chi_{i,t} \geq 0 \quad \perp \quad CAP_{i} \geq 0 \qquad \forall i > 0 \qquad (2.110)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (2.111)$$

$$1 - u^{e}_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (2.112)$$

Selecting Critical Utilisation Levels

$$1 - \psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.113)$$

$$\psi^{0}_{n,t} \sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC_{j,t}} + \psi^{3}_{n,t} z_{j,n\text{-}1,t} + 2\psi^{4}_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j,n > 0,t \qquad (2.114)$$

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t} \sum_{i} z_{i,n\text{-}1,t} u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.115)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.116)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC_{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC_{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.117)$$

$$-\sum_{j} z_{j,n\text{-}1,t} z_{j,n,t} + \psi^{5}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.118)$$

$$\sum_{j} z_{j,n,t}{}^{2} - 1 \geq 0 \quad \perp \quad \psi^{4}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.119)$$

$$1 - u^{e}_{n,t} \geq 0 \quad \perp \quad \psi^{5}_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (2.120)$$

Ordering Utilisation Levels

$$-\mathrm{r.u_{k,t}} + \phi^{0}_{r,t} + \phi^{f}_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k,r,t \qquad (2.121)$$

$$-\mathrm{r}.u^{e}_{n,t} + \phi^{0}_{r,t} + \phi^{e}_{n,t} \geq 0 \quad \perp \quad x^{e}_{n,r,t} \geq 0 \qquad \forall n,r,t \qquad (2.122)$$

$$1 - \sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{0}_{r,t} \geq 0 \qquad \forall r,t \qquad (2.123)$$

$$1 - \sum_{r} x_{k,r,t} \geq 0 \quad \perp \quad \phi^{f}_{k,t} \geq 0 \qquad \forall k,t \qquad (2.124)$$

$$1 - \sum_{r} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{e}_{n,t} \geq 0 \qquad \forall n,t \qquad (2.125)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_{k} x_{k,r,t}\mathrm{u_{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} = 0 \qquad \forall r,t \qquad (2.126)$$

$$L_{r,t} - \sum_{k} x_{k,r,t}\mathrm{L_{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r,t \qquad (2.127)$$

$$L^{e}_{n,t} - \mathrm{L}_{0,t} + \sum_{k} \frac{\mathrm{L}_{k-1,t} - \mathrm{L}_{k,t}}{\mathrm{u}_{k,t} - \mathrm{u}_{k-1,t}} u^{part}_{k,n,t} = 0 \qquad \forall n,t \qquad (2.128)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n,t \qquad (2.129)$$

$$\mathrm{u}_{k,t} - \mathrm{u}_{k-1,t} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < k < K, n$$

Initial Conditions:

$$u^{e}_{0,t} = 0 \qquad \forall t \qquad (2.131)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (2.132)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (2.133)$$

While the complementarity structure is useful for theoretical analysis, it is not, without adjustment, always the best solution approach. Nevertheless, despite the many alternative solution processes, complementarity theory provides the only consistent framework we are aware of, which is capable of melding and encoding both algorithmic and optimisation approaches, as well as providing an, albeit computationally burdensome, pathway to solution with current solvers.

### 2.7.2 Nested Solution

As shown by the algorithm in Section 2.4.1, in its most basic form the problem can be viewed as a nested problem. The first optimal trade-off is with shortage, and in this simplified structure no technology operating below the peaker in the merit order can influence that trade-off. The solution approach is therefore to determine the optimal peaking technology, along with the optimal trade-off overall, as a function of sub-period performance. This defines the profitability of the peaker of last resort, and forms a fixed portion of the profit or cost recovery available to other technologies.

We have the following inequality describing the technology i* that is best placed to generate with a utilisation level of $u_1$, then:

$$\lambda_1 = \mathrm{MC}_{i^*} + \frac{\mathrm{FC}_{i^*}}{\mathrm{u}_1} \qquad (2.134)$$

An energy price higher than this level will conflict with the equilibrium condition (2.20). This price may correspond to a shortage cost, for which the fixed cost is typically assumed to be zero, or it may be a peaking technology with sufficiently low marginal cost and capital recovery factor that it is a more economical alternative at $u_1$. Whichever technology is the least cost, and whether or not two or more technologies are tied, the price, $\lambda_1$, will be unique.

We now consider which technology is the least cost technology to generate with the next utilisation level, $\mathrm{u}_2$. Following the same approach as above, at $\mathrm{u}_2$, we have:

$$FC_j \geq (\lambda_1 - MC_{i*})u_1 + (\lambda_2 - MC_{j*})(u_2 - u_1) \qquad \forall i > 0 \qquad (2.135)$$

(2.135) holds for all technologies, and with strict equality for the marginal technology, j*. Recognising the first term $(\lambda_1 - MC_{i*})u_1 = FC_{i*}$, we have the following expression for $\lambda_2$, the energy price at $u_2$ :

$$\lambda_2 = MC_{j*} + \frac{(FC_{j*} - FC_{i*})}{(u_2 - u_1)} \qquad (2.136)$$

This is also uniquely defined by the cost parameters of each technology and the utilisation levels used to represent the LDC.

By an induction argument, the PDC is uniquely defined, as the expression relating fixed costs results in selecting the next technology in the merit order as a function of the previous optimal choice. At each stage the PDC and the technology most efficient at the utilisation level under consideration is determined. By progressing through each utilisation level, we can determine the solution in a nested fashion. Unfortunately, the approach taken here is not robust in situations where technological selections lower in the merit order influence those above. In these cases, the algorithm would require additional macro iterations to develop a solution in which high level trade-offs and choices are consistent with lower level choices and trade-offs.

### 2.7.3 Decomposition

An alternative, and more general approach, involves dividing the problem between the basic investment and market clearance (economic) problem and the utilisation level determination problem. The former includes the standard economic complementarity constraints governing market clearing and investment, while the latter includes the complementarity constraints that describe the definition, ordering and calculation of utilisation levels designed to optimise the representation of the problem.

The decomposition is seeded with an initial solution to the economic problem using only the load and utilisation levels that describe the LDC. This generates the dual variables that measure marginal capacity values and marginal fuel values that determine the optimal technological trade-offs. The second stage encompasses the definition of critical optimal trade-offs, the ordering of these trade-offs, and the load levels corresponding to those trade-offs. The process then reverts to the initial problem, but this time operating on the newly defined LDC. That problem is re-solved and generates a new set of marginal capacity and fuel values. The process continues until convergence.

This separation of the problem in this form enables separate choices regarding which of the modelling paradigms, such as optimisation, complementarity or algorithms, are most appropriate at each level. For example, in simple cases, a fast algorithm could develop the system representation more quickly than a complementarity formulation reconstructing the screening curve lower envelope. In other cases, the simple algorithm may require modification or an additional level of iteration to achieve this and may be less effective than a direct approach. Further, in cases of perfect competition, a direct optimisation could be used with the correct set of utilisation levels, whereas when the market is gamed a complementarity formulation may be desirable. Our approach is general, in that whether

algorithm, optimisation or complementarity formulations are invoked, they can all be unified if desired in a single conceptual framework, whether or not that is the desired solution technique.

### 2.7.4 Existence and Uniqueness

The theory of existence and uniqueness of complementarity problems is well surveyed in Harker & Pang (1990).

The existence of a solution to our problem is clearest when viewed in the form of a decomposition. The lower level problem determines a system representation by selecting and ordering appropriate utilisation levels as defined by the higher-level problem. The higher-level problem has a solution as was shown in Section 1.5.3, that is independent of the particular specification of the LDC.

Uniqueness is a more elusive solution property, and technically our framework does not provide a unique solution. The complementarity problem we have formulated is known to have multiple solutions, and therefore a priori we know it will not be susceptible to uniqueness analysis via complementarity theory. Nevertheless it is also the case that, while not conforming to the uniqueness results defined in complementarity theory, if we modify our definition of a unique solution to something more practical, such as a unique set of capacity investment, generation and market clearance variables, then a "practically" unique solution property can be demonstrated.

We can further advance the discussion of practical uniqueness by examining the problem in the context of a decomposition. Given a capacity stock and a system representation, the market clearance is unique. The uniqueness of the market clearance implies that the dual variables that define the marginal profitability of each technology are also unique. In turn, technological profitability uniquely defines both the marginal value of additional capacity and the optimal trade-off between technologies, which is the system representation. Not all system representations and capacity stocks will correspond to an equilibrium of the system. Starting from an initial point, the system will incentivise alterations in capacity levels and the utilisation levels that correspond to optimal trade-offs and define the system representation. Both adjustments are monotonic in capacity so that, as capacity of a technology is added, the earnings of that technology shrink while the marginal utilisation range of the technology expands. As these adjustments occur in unison the system will approach a fixed point where there is no incentive to adjust capacity values and/or the system representation.

## *2.8 Summary & Conclusions*

In this chapter, we have resolved several issues with the conventional optimisation approach. Those issues are related to the choice of the LDC representation, the implicit restriction on generation functions and the use of exogenous utilisation levels.

Following the lead of Chapter 1, we begin by directly comparing screening curve analysis with conventional optimisation formulations. We identify the relative endogeneity of utilisation levels and the compact definition of the PDC in screening curve analysis. We also note that one alternative to the removal of the implicit restrictions within conventional formulations, is to minimise the influence of them, and we describe how increasing the granularity of the model both shrinks the size and the duration, in utilisation terms, of the error. The cost of such an approach may be small in the context of

linear programming, but the same brute force approach would be less attractive in cases where convex optimisation was not capable of representing the model structure. In any case, a superior solution is to remove the effect of the restriction, as opposed to the restriction itself, altogether by combining endogenous utilisation levels with the fixed utilisation levels in the optimisation.

The broad strategy for achieving this integration is as follows:

- Determine utilisation corresponding to optimal trade-offs
- Prune the set of optimal trade-offs to those that are relevant
- Integrate this set of utilisation levels with the exogenous utilisation levels that describe the LDC
- Define load levels corresponding to the endogenous utilisation levels.

In effect this approach optimises the otherwise sub-optimal conventional optimisation formulation. While this is conceptually an optimisation, without optimisation of the formulation, the conventional approach can only approximate the solution to the problem. An adaptation of the conventional optimisation may be possible, but ultimately any adaptation must address the same issues listed immediately above so it is not obvious what form that optimisation would take. In particular, any adaptation must address the defining feature of the problem, which is the endogenous optimal trade-offs which allow for the definition of optimal generation functions. So while the problem can be conceptually viewed as an optimisation, and with the exception of a technical optimisation based on the complementarity pairs we have identified, we are unaware of any optimisation that can address this general problem.

We begin with a set of market clearances, and then develop an investment constraint. We then introduce sub-periods and, in doing so, cross a modelling threshold, beyond which the simple screening curve diagram can no longer represent the global solution. In this environment, technologies are selected based on their suitability across a range of conditions, perhaps without being absolutely suited to any one. The extension of the model to address sub-periods requires the specific definition of sub-period profitability. This enables the precise definition of imputed marginal capacity values, which can then be used in an analogous fashion to calculate the critical utilisation levels in each sub-period. In turn, these utilisation levels precisely define the marginal operating ranges of each technology in each sub-period as well as the sub-period PDC.

Having demonstrated the nature of the solution, Sections 2.4 develops the necessary complementarity conditions to define the critical utilisation levels. In the simplest single period deterministic example, critical utilisation levels can be determined a priori and inserted as data. In more complex cases, the critical system utilisation levels are not able to be determined a priori and we provide a set of complementarity conditions for defining the relevant optimal trade-offs. Where the merit order is not consistent we must define optimal trade-offs in both directions. This generally results in the determination of one interior solution, where the incentive is to adjust utilisation to remove the cost differential between two technologies, and one boundary solution in which the limits of utilisation are invoked in the complementarity condition.

That process generates redundant utilisation levels, but even in the case where the merit order is fixed we are still only interested in those critical utilisation levels that define operational trade-offs

between marginal technologies and in doing so define the PDC. In section 2.5, we reduce the number of utilisation levels under consideration to a smaller set using a simple approach that culls all but one optimal trade-off for each technology. A more advanced, although not necessarily more efficient approach is a systematic search along the screening curve envelope. Sub-period trade-offs are ultimately based on global capacity choices and the nature of the LDC.

These critical utilisation levels are then combined with exogenous utilisation levels, using a ranking optimisation in complementarity form. The ranking optimisation extends beyond the most basic implementation of its type to address the joint sorting problem that exists as a result of considering an exogenous and endogenous set of utilisation levels. Once sorted, the load levels corresponding to the pruned and sorted set of utilisation levels are endogenously determined by a set of complementarity conditions designed with a piecewise, and generally non-convex, LDC in mind.

Finally, we briefly discuss solution methodology and the existence and uniqueness of the solution. As noted, the complementarity formulation is not the only solution method. Other algorithms are no doubt possible, and could be investigated, but in this case the complementarity framework provides a succinct view of all the incentives and constraints of interest to investors and operators. In addition, it permits simulation of the motivation of notional or meta-entities, which are invoked to define the most useful representation of the problem from a range of possible representations, each of which can produce a solution. Furthermore, complementarity theory offers a potentially unifying statement of the problem that will also be amenable to the consideration of strategic situations.

In return for accepting the additional computational burden, the inclusion of endogenous utilisation levels corresponding to optimal trade-offs:

- Clearly defines the optimal utilisation ranges for each technology;
- Enables generation and investment decisions to be considered simultaneously
- Correctly identifies market prices consistent with market clearances, and therefore does not contaminate any further analysis reliant on price response;
- Avoids the inclusion of an unknown, but large, number of fixed utilisation points, which must be sufficient to represent the solution, and provide useful sensitivity analysis
- Enables a clearer assessment of risk; and
- Forces a conceptual focus on what is at the heart of investment and/or optimal plant mix problems: The optimal trade-off between technologies.

As far as we are aware, our approach is novel and consists of a set of complementarity conditions that express the inherent logic of screening curves, along with the traditional principles of market clearing and investment.

# 3 TECHNOLOGICAL CONSIDERATIONS

## *3.1 Introduction*

It is important to illustrate that the fundamental approach taken in Chapter 2 is reasonably extensible, and this is the goal of Chapters 3,4 and 5. A common, and valid, criticism of screening curve analysis is that relatively minor extensions in complexity make it intractable. However, the logic of screening curve analysis remains valid and we use it as we consider some technological issues that require extension of the complementarity model in Chapter 2. We do so to illustrate how these might be implemented, whether the conditions for determination of endogenous utilisation levels require adjustment and, in the case adjustment is required, what form that adjustment takes.

In this chapter, we further generalise the technological generation cost structure in several ways. We begin by generalising the previous linear cost function to a convex piecewise linear cost function. While not novel, we discuss the relevance of non-convexities in cost functions, and the reasonableness of assuming these do not exist in our framework. We then develop a formulation treating investment in such technologies as a parcel of technologies bound by a proportionality constraint. Finally, we then examine how complementarity conditions relating to investment and the determination of endogenous utilisation levels are affected.

In the Sections 3.3 and 3.4 of this chapter we discuss technological limitations. Beginning with capacity limitations, we investigate how these can be incorporated into our framework. We extend this to include a stylised model of mothballing and reinstatement. We then introduce energy limits and generalise these by sub-period to illustrate how storage restrictions can be included in the framework.

In Section 3.5 we introduce an entirely novel approach that defines ranges of configurable technologies in the screening curve context. In principle, the introduction of configuration adds a further dimension to the investment problem. Using an assumed quadratic trade off in the cost structure, we develop an optimised piece-wise quadratic total cost function that, despite requiring a different interpretation to a standard cost function, enables the representation of optimised technological configurations. The introduction of a piece-wise quadratic cost structure into screening curve analysis has implications for the number of trade-offs between technologies and the means of integration with linear technologies. To determine endogenous utilisation levels, we address the algorithmic requirements of avoiding imaginary roots. Finally, we re-define the investment condition to reflect a PDC comprised of piecewise linear and constant sections.

## *3.2 Generalised Cost Structures*

### 3.2.1 Piecewise Constant Marginal Costs

Thus far we have assumed that marginal costs are constant and therefore there are no economies, or diseconomies, of scale in generation. As shown by way of example in Figure 22, generation technologies may have start-up costs and efficient operating zones, which can be interpreted as economies or diseconomies of scale, so that as shown in Figure 22, based on economic incentives the

various operating tranches would not be selected in a logically consistent order. Minimum operating levels can also be viewed in this framework as an infinite cost at capacity levels below the minimum operating level.

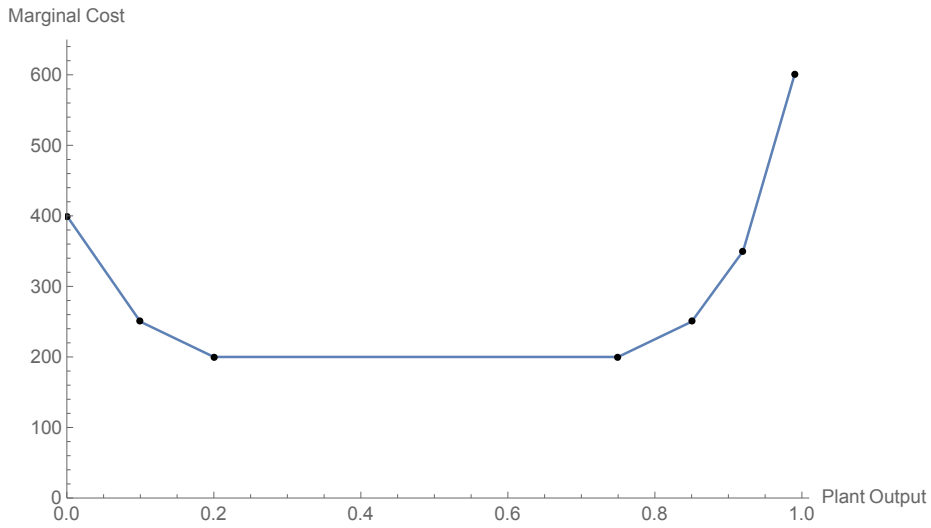



**Figure 22: Indicative Cost Structures**

At the technological level, as opposed to the plant level, the assumption that there are no minimum operating levels will not bias the analysis significantly. While any minimum operating levels may be based on physical requirements, for any reasonably sized system the minimum operating level of a single plant is scarcely significant when modelling at a technological level. Similarly, when considered at the technological level, inefficient operating zones are unlikely to feature among equilibrium outcomes, as generators will quickly approximate efficient operation using a subset of their own available plant to best capture efficient operating ranges. This provides some comfort that an analysis based on the assumption that costs don't exhibit such features will be accurate even where those characteristics are not modelled explicitly.

The implication of the second assumption, that there are no diseconomies of scale, does introduce systematic bias. Specifically, this assumption implies that generators can use all of the available capacity at an efficient cost. Ultimately though, the efficient generation range of all available plants of technology i is exhausted. Further output, up to the capacity rating of the plant, may be relatively inefficient or result in additional maintenance and repair costs if the lifespan of the plant is to be maintained. Ignoring this phenomenon systematically misstates technological profitability to a degree that is dependent on the extent, and the range over which, plant efficiency falls. Unlike the issue of increasing returns to scale, the issue is not resolved, or diminished, by adopting a large market assumption.

The cost structure of each technology contemplated so far consists of a constant marginal cost per unit of utilisation and a fixed cost, giving a linear total generation cost in terms of utilisation. We now relax this assumption to consider a piecewise linear total generation cost, which in turn implies the consideration of step-wise marginal cost curves in which the steps represent several discrete marginal

cost ranges for each technology i. To formulate a piecewise linear total cost function in a way that is consistent with the overall concept of determining optimal utilisation levels we no longer consider technology i as a single technology, but as a group of linked technologies, associated with i, for which the fixed costs are incurred collectively and for which the individual capacity of each related technology is a function of the installed capacity of technology i and a technological assessment of the costs of operating that technology in various modes, i(m).

We assume there are M operating modes, indexed by m=1…,M, for each technology so that i(m) denotes technology i operating in mode m, and corresponds to a single cost tranche. $FC_i$ continues to represent the fixed cost of a unit of capacity of technology i, however that unit of capacity is broken into M portions, with $MC_{i(m)}$ being the marginal cost of technology i operating in mode i(m).

For their investment, investors effectively receive a portfolio of notional technologies, some of which may or may not be profitable, with portfolio share, $\alpha_{i(m)}$, defined according to engineering data on technological capabilities to be the proportion of capacity of technology i that is applicable to the operating range m for technology i:

$$CAP_{i(m)} - \alpha_{i(m)} CAP_i = 0 \quad \perp \quad \chi_{i(m)} \; free \qquad \forall\, i(m)\big|_{i>0} \tag{3.1}$$

The equilibrium investment condition records sub-period earnings as the sum of earnings for each operating mode of technology I in sub-period t, weighted by the fixed proportions that define the cost structure. Total earnings are found by taking the sum of sub-period earnings weighted by the duration of the sub-period:

$$FC_i - \sum_t w_t \sum_{i(m)} \alpha_{i(m)} \chi_{i(m),t} \geq 0 \quad \perp \quad INV_i \geq 0 \qquad \forall i > 0 \tag{3.2}$$

Where fixed costs exceed the marginal profitability of additional capacity of technology i, then $INV_i = 0$ in equilibrium. Where the opposite is true, investment will be incentivised and continue until parity is achieved between fixed costs and the marginal profitability of technology i.

Without (3.1), the inefficient operating modes of a particular technology would be dominated by the efficient modes as capacity of either has an identical fixed cost. The imputed value of each technological mode, $\chi_{i(m)}$, adjusts to the point at which the proportionality constraint (3.1) is obeyed. This adjustment requires the upward movement of the total cost curve associated with more efficient operating modes, reflecting an increase in the marginal capacity value associated with that operating mode. For similar reasons, we also observe downward movement of the cost curve associated with less efficient operating modes, reflecting a decrease in the marginal capacity value associated with that operating mode. The adjustment ceases when either the proportionality constraint is satisfied, or when the overall technology ceases to be profitable. The equilibrium setting is one in which the value of relatively efficient capacity is higher than its installation cost, and the value of relatively inefficient capacity is lower than installation cost. It is not necessarily the case that investment is justified at all, and where the weighted return of all operating modes is negative there will be no capacity built.

As each technological operating mode operates with a different marginal cost and we have specifically ruled out economies of scale at the technology level then, subject to observing the offer rules regarding the number of tranches available to be specified by each plant in a competitive offer, they will be offered to the market as individual technologies. Accordingly, the market clearing conditions require generalisation to refer to operating modes i(m), rather than actual technologies.

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i(m),t}} + \varphi^{+}_{i(m),r,t}\Big|_{i>0} \geq 0 \quad \perp \quad GEN_{i(m),r,t} \geq 0 \qquad \forall i(m), r, t \tag{3.3}$$

$$\sum_{i(m)} GEN_{i(m),r,t} - \mathrm{L}_{r,t} = 0 \quad \perp \quad \lambda_{r,t} \; free \qquad \forall r, t \tag{3.4}$$

$$CAP_{i(m),t} - GEN_{i(m),r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i(m),r,t} \geq 0 \qquad \forall\, i(m)\big|_{i>0}, r, t \tag{3.5}$$

The definition of sub-period earnings remains the same but is generalised to include all operating modes as if they were separate technologies:

$$\chi_{i(m),t} - \sum_{r<R} \varphi^{+}_{i(m),r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i(m),t} \; free \qquad \forall\, i(m)\big|_{i>0}, t \tag{3.6}$$

Finally, the condition defining optimal trade-offs must also be generalised:

$$\chi_{i(m),t} - \chi_{j(m),t} - (\mathrm{MC}_{\mathrm{j,t}} - \mathrm{MC}_{\mathrm{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i(m), j(m) \neq i(m), t \tag{3.7}$$

For each sub-period, the optimal trade-off is evaluated according to the available capacity of each technology and is defined by the interaction between sub-period profitability and the marginal cost of each technology. Therefore, the optimal trade-off between operating modes of the same technology is a valid optimal trade-off. The capacity of each operating mode is required to be consistent with the technological cost curve and is defined in (3.1)

## *3.3 Capacity Inflexibility*

### 3.3.1 Introduction

In the general framework, we have already dealt with capacity inflexibility in the form of ordinary capacity, which cannot respond to:

- Variations in load within a sub-period, as represented by sub-period LDC's;
- Variations in load between sub-periods; or
- Cost and other variations between sub-periods.

The consideration of additional capacity inflexibility may be motivated by issues such as the limitation of opportunities, or the availability of pre-existing capacity, for which the decision is to operate, retire or mothball. Whatever the reason for the inflexibility, the cost of inflexibility or the value of a limited opportunity should be assessed in terms of market performance and returns. These can then be compared to the actual cost of creating and/or maintaining that capacity. In the case of an investment decision, the relevant cost would be the total of construction/installation costs and fixed operating and maintenance costs, while with a retirement or mothballing decision the firm would wish to ascertain

whether fixed operating and maintenance costs are able to be covered, as capacity installation costs are sunk.

### 3.3.2 Opportunity Limited Technologies

Opportunity limits arise as a result of the availability of a specific development site, permit or load response option. In these cases, the installation options for a particular technology are restricted to a limited range of options, each of which is specified by the quantum of capacity available and a cost per unit. Where the issue is not global as much as it is site or permit specific, then the representation of unique opportunities, such as hydro or wind generation opportunities, that are of significant scale relative to the market size may necessitate the description of each such opportunity as individual technologies, for which design limitations apply.

Our discussion primarily considers global limits at the technological rather than plant level, and Figure 23 shows a global limit of the capacity of a single technology. As shown, when a capacity limit is introduced for consideration, capacity and the optimal utilisation of the technology are both (weakly) lower than they would be in an unconstrained equilibrium. At the same time, incentives exist for increasing the capacity of adjacent technologies to compensate for the limitation on technology i. Graphically, it is apparent that at that capacity limit, the marginal earnings and the imputed valuation of an additional unit of capacity are higher than in the unrestricted equilibrium, and therefore must also exceed $\mathrm{FC_i}$, the fixed cost of capacity.

If we consider the case presented in Figure 23, the optimal trade-offs between a technology i and the adjacent technologies i+ and i- may be interpreted as providing the solution to the following set of equations:

$$CAP_i = \left( \frac{\mathrm{L_{k+1}} - \mathrm{L_k}}{\mathrm{u_{k+1}} - \mathrm{u_k}} \right) \left( u^e_{n+1} - u^e_n \right) \tag{3.8}$$

$$u^e_{n+1} = \frac{\mathrm{FC_{i+}} - \chi^+_i}{\mathrm{MC_i} - \mathrm{MC_{i+}}} \tag{3.9}$$

$$u^e_n = \frac{\chi^+_i - \mathrm{FC_{i-}}}{\mathrm{MC_{i-}} - \mathrm{MC_i}} \tag{3.10}$$

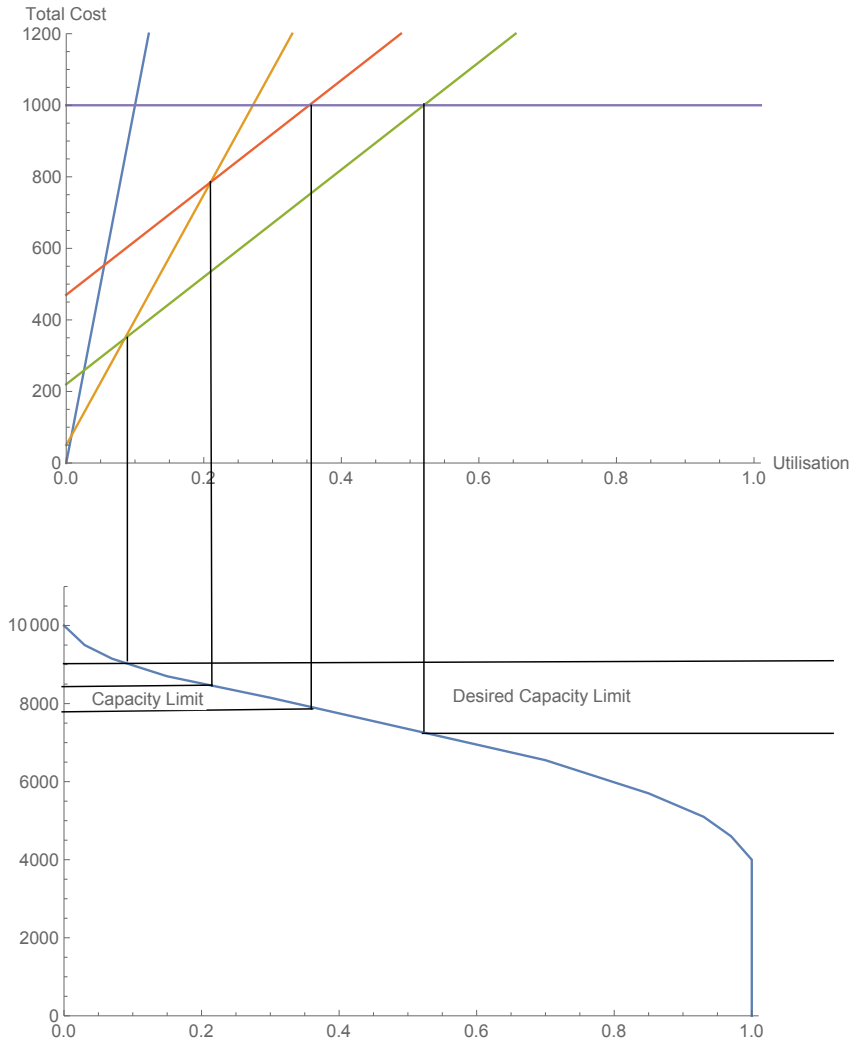



**Figure 23: Capacity Limited Technologies**

The mathematics confirms the graphical relationship. As the optimal utilisation range for technology i shrinks along with the capacity level, the imputed marginal value of capacity of technology i, $\chi_i^+$, rises. This process continues until the value of capacity has adjusted sufficiently to incentivise a capacity level that respects the capacity limit in (3.8).

To introduce this structure into a complementarity formulation, we begin with a complementarity condition to restrict capacity to a maximum level, $\mathrm{CAP}_\mathrm{i}^+$:

$$\mathrm{CAP}_\mathrm{i}^+ - CAP_i \geq 0 \quad \perp \quad \chi_i^+ \geq 0 \qquad \forall i \tag{3.11}$$

The complementarity variable $\chi_i^+$ represents the value of a relaxation in the capacity restriction or, alternatively, the value to the system of additional capacity when capacity is at its predefined maximum. When the optimal capacity choice is less than the pre-defined maximum, the constraint is not binding and $\chi_i^+ = 0$.

Barring the opportunity limit, if the capacity of technology i was below its optimal value, then $\mathrm{FC_i} < \sum_t \mathrm{w_t}\chi_{i,t}$, where $\chi_{i,t}$ continues to represent the imputed marginal value of capacity in each sub-period and, where appropriate, adjusts to reflect the higher equilibrium returns that flow from capacity restricted to a maximum level. Normally, investment in such a technology would occur, until the marginal benefit of further investment was equal to fixed costs of doing so. But no such option exists when $CAP_i = \mathrm{CAP_i^+}$. $\chi_i^+$ measures the marginal cost to the system of the requirement to maintain the capacity of a particular technology at a level no greater than $\mathrm{CAP_i^+}$. We know from a basic optimisation of the form struck in Chapter 1, that the addition of a capacity constraint would result in the first order condition with respect to capacity gaining an additional term. Accordingly, we introduce $\chi_i^+$ into the equilibrium investment condition, where it is quantified as the difference between the fixed cost and the weighted marginal profitability of technology i. The equilibrium investment condition now reflects the possibility that constraints on capacity restrict complete adjustment and permit individual technologies to be marginally profitable in equilibrium.

$$\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} + \chi_i^+ \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0, t \tag{3.12}$$

Although the actual optimal trade-offs between technologies will adjust until the capacity limit is respected, the impact of the capacity limit is the naturally accounted for in $\chi_{i,t}$ so that the complementarity condition that defines optimal trade-offs still does so and requires no further adjustment:

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^e_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^e_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \tag{3.13}$$

A similar analysis is available for any technology with a mandated minimum capacity:

$$\mathrm{CAP_i^-} - CAP_i \geq 0 \quad \perp \quad \chi_i^- \geq 0 \qquad \forall i \tag{3.14}$$

If the capacity of technology i was required to be above its optimal value, then $\mathrm{FC_i} > \sum_t \mathrm{w_t}\chi_{i,t}$. This would reflect over-investment in technology i. Normally, investment in technology i would be scaled down or existing capacity would be retired, until either the marginal benefit of further investment was equal to fixed costs of doing so, or capacity reached zero, the de-facto minimum. But no such option exists when $CAP_i = \mathrm{CAP_i^-}$. $\chi_i^-$ measures the marginal cost to the system of the requirement to maintain the capacity of a particular technology at a level no less than $\mathrm{CAP_i^-}$. We now introduce $\chi_i^-$ into the equilibrium investment condition, where it is quantified as the difference between the fixed cost and the weighted marginal profitability of technology i. The equilibrium investment condition now reflects the possibility that constraints on capacity may restrict full equilibration of capacity in either direction and through the inclusion of additional variables tied to complementarity conditions this condition now permits individual technologies to be marginally profitable or unprofitable in equilibrium when capacity constraints are binding.

$$\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} + \chi_i^+ - \chi_i^- \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0, t \qquad (3.15)$$

The marginal value of the opportunity or restriction is given by $\chi_i^+ - \chi_i^-$. At most one of these terms will be positive, depending on whether the capacity maximum or minimum is binding. We have described opportunity limits in terms of their cost to the system, however we note the owner or controller of the opportunity or the restriction, such as site owner or license owner, is effectively the owner of those constraints, whether they represent a right or obligation. If the constraint could be traded in a perfectly competitive, and linear, secondary market, the effective owner of the constraint should be willing to receive or pay the associated rental in the amount of $\chi_i^+ - \chi_i^-$ for each unit of the restriction traded.

*Discrete Investment Options*

Although we are considering restrictions on the range of available capacity options for opportunity limited technologies, that restriction is being applied at a technological level, rather than plant level. In a large market, aggregation to the technological level makes the assumption of continuous investment more reasonable, although opportunity limits are often associated with unique or rare opportunities that cannot be easily replicated or scaled. For example, opportunities for investment in technologies such as hydroelectric generation are significantly restricted by the natural characteristics of the installation location. In these cases, it may not be possible to build the plant size that is optimally indicated by the screening curve analysis. Instead the capacity choice may be restricted to a set of possible installation sizes governed by engineering or other practical considerations.

From a mathematical perspective, the important feature of this situation is that the options are discrete and, for example, include the option to not invest at all. An equilibrium representation of the decision to invest with discrete development opportunities requires consideration of non-convexities. In particular, the decision of whether to invest in a discrete quantity of a particular technology is one in which whatever option is chosen, it may appear to be the wrong decision ex post. If the investment progresses, it may cannibalise the very opportunity it sought to capitalise on. Yet, if it does not, there remains an apparent opportunity for investment or entry. The ideal capacity level is that which captures sufficient returns to recover its capital cost but is not overbuilt and therefore cannibalising its own income stream. This fine-tuning is not possible when only discrete capacity choices are available.

To evaluate such an opportunity (or obligation), we can reverse our analysis and calculate the value of the option for any fixed MW quantity of plant to be built by considering the discrete capacity choices through setting the minimum and maximum capacity limits to the desired capacity setting. Unfortunately, this procedure must be considered in iterative fashion and separate from the equilibrium determination of the complementarity problem.

### 3.3.3 Existing Capacity

Existing capacity is perhaps the most prevalent form of limitation on the capital stock. The equilibrium level of existing capacity is itself dependent on the equilibrium timeframe under consideration. Existing capacity does not exist when equilibrium is considered in its ultimate sense. But investors are

fundamentally interested in the lifespan of their investment so that, in anything other than the long-term concept of equilibrium, existing capacity can be assumed to play some part in the operation of the market, and should be significant to investors. Although the declining performance or outmoded specifications of older plant relative to new installations of similar technologies suggest a changing role for aging plants, the effect of technological progress is somewhat muted by the sunk costs of existing capacity which make the economic justification for inclusion of pre-existing plant in an optimal plant mix relatively more favourable than it is for new investment, ceteris paribus.

We begin by considering the impact of existing capacity without possibility of retirement or mothballing. We can incorporate existing capacity, $\mathrm{CAP}_i^0$, in the same fashion as the minimum capacity constraint of the previous section, noting that $\mathrm{CAP}_i^0$ is to some degree parameterised by the timeframe underlying the equilibrium analysis:

$$CAP_i - \mathrm{CAP}_i^0 \geq 0 \quad \perp \quad \chi_i^0 \geq 0 \qquad \forall i \tag{3.16}$$

in the absence of retirement or mothballing, one or other minimum capacity constraint will be redundant. We include both as it is not appropriate to conflate the two constraints, particularly if post-solution analysis might consider different policy settings, or equilibrium timeframes, thereby potentially switching the active constraint. Because (3.16) is of the same form as the minimum capacity constraint, so too is the form of the required alteration to the equilibrium capacity condition:

$$\mathrm{FC}_i - \sum_t \mathrm{w}_t \chi_{i,t} + \left(\chi_i^+ - \chi_i^- - \chi_i^0\right) \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0, t \tag{3.17}$$

*Retiring Capacity*

Capacity and investment are synonymous whenever the timeframe of the equilibrium under consideration extends beyond the remaining lifespan of the technology concerned. However, in shorter timeframes, existing capacity may be relevant. We consider shorter timeframes because investments themselves have a finite lifespan. Accompanying this reality is the possibility that retirement is an economic option, available to the owners of existing plant. This option affords the level of capacity some downward flexibility. To that end, we separate the capacity decision into its composite parts; an investment choice, and a retirement choice.

We assume that the cost of retirement will be incurred whether the plant is retired for economic or age-related reasons, and that the cost of doing so is invariant, in discounted terms, to either the timeframe under consideration or the age of the plant when retirement occurs, so that aside from spot market returns, no other benefits or costs are relevant. With existing capacity, the fixed costs of installation are sunk. Therefore, ignoring other benefits and costs, to retain the capacity the owner of existing capacity need only ask whether the capacity concerned will at least contribute to the recovery of sunk costs, and not whether it will achieve full recovery of those costs.

Fixed operating costs, $\mathrm{FOC}_i$, are those fixed costs that relate to the operation of the plant whether or not electricity is being generated. Existing capacity can contribute to sunk cost recovery, and is therefore worth retaining, whenever:

$$\sum_t \mathrm{w}_t \chi_{i,t} - \mathrm{FOC}_\mathrm{i} \geq 0 \qquad \forall i > 0, t \qquad (3.18)$$

Conversely, when (3.18) does not hold, existing capacity does not cover fixed operating costs and rather than recovering some portion of sunk costs, augments them with further operating losses. In that case, some capacity should be retired. As capacity is retired, the profitability of remaining capacity increases, until eventually earnings rise to the level where fixed operating costs are being covered. The following complementarity condition reflects this logic in the form of an equilibrium retirement condition:

$$\sum_t \mathrm{w}_t \chi_{i,t} - \mathrm{FOC}_\mathrm{i} \geq 0 \quad \perp \quad RET_i \geq 0 \qquad \forall i > 0, t \qquad (3.19)$$

To ensure capacity remains non-negative, retirement is limited to the level of initial capacity available:

$$\mathrm{CAP}_\mathrm{i}^0 - RET_i \geq 0 \quad \perp \quad \chi_i^{RET} \geq 0 \qquad \forall i \qquad (3.20)$$

If existing capacity is only partially retired then $\chi_i^{RET} = 0$, implying that, through retirement, capacity has adjusted until the marginal profitability of the capacity is equal to the fixed operating costs of retaining and operating that capacity. Where existing capacity is fully retired, $\chi_i^{RET} \geq 0$. At this point, it is possible that the LHS condition of (3.19) does not hold, indicating a desire for further retirement which, from (3.20) is not feasible. Accordingly, (3.19) requires modification to reflect the limitations on capacity retirement decisions that restrict the full equilibration of capacity levels, in this case, to negative levels.

$$\sum_t \mathrm{w}_t \chi_{i,t} - \mathrm{FOC}_\mathrm{i} + \chi_i^{RET} \geq 0 \quad \perp \quad RET_i \geq 0 \qquad \forall i > 0, t \qquad (3.21)$$

Where $\mathrm{FOC}_\mathrm{i}$ exceeds the profitability from operating technology i, and all existing capacity has been retired we have $RET_i > 0$ and $\chi_i^{RET} = \mathrm{FOC}_\mathrm{i} - \sum_t \mathrm{w}_t \chi_{i,t} > 0$, which could be interpreted as the subsidy necessary to retain the last unit of unretired capacity of technology i.

Ignoring capacity limitations, we replace the equilibrium capacity condition (2.110), with the equilibrium investment condition (3.22), which restates the equilibrium capacity condition in terms of a new investment variable, $INV_i$.

$$\mathrm{FC}_\mathrm{i} - \sum_t \mathrm{w}_t \chi_{i,t} \geq 0 \quad \perp \quad INV_i \geq 0 \qquad \forall i > 0 \qquad (3.22)$$

If $INV_i > 0$ and $RET_i > 0$, then from (3.22) and (3.21) we have

$$\mathrm{FC}_\mathrm{i} - \sum_t \mathrm{w}_t \chi_{i,t} + \sum_t \mathrm{w}_t \chi_{i,t} - \mathrm{FOC}_\mathrm{i} + \chi_i^{RET} = 0 \qquad \forall i > 0 \qquad (3.23)$$

This simplifies to:

$$\mathrm{FC}_\mathrm{i} - \mathrm{FOC}_\mathrm{i} + \chi_i^{RET} = 0 \qquad \forall i > 0 \qquad (3.24)$$

Equation (3.24) is a contradiction, as $\chi_i^{RET} \geq 0$ and $\mathrm{FC_i} - \mathrm{FOC_i} > 0$. Unsurprisingly, there is no value in investing in capacity, while simultaneously retiring capacity of the same. Instead, when we consider the criteria for new investment alongside the criteria for retirement, we find a profitability range in which fixed operating costs are being covered but full cost recovery is not being achieved and, as a result, neither investment nor decommissioning will occur:

$$\mathrm{FOC_i} \leq \sum_t \mathrm{w_t} \chi_{i,t} \leq \mathrm{FC_i} \qquad \forall i > 0 \qquad (3.25)$$

Where prices are such that investment in technology i is justified, investment will occur until prices and the marginal profitability of technology i fall to equate to $\mathrm{FC_i}$, at which point equilibrium will have been attained. Where prices fall short of justifying investment, but exceed $\mathrm{FOC_i}$, existing capacity will be retained, with no retirements or further investment. Where prices fall short of $\mathrm{FOC_i}$, retirement is incentivised. As retirement is actioned and capacity falls, either retirement is partial and profitability increases until the equilibrium returns precisely cover $\mathrm{FOC_i}$, or all existing capacity will be retired and $\mathrm{FOC_i}$ will still exceed the marginal profitability for technology i.

Viewing the situation dynamically, we see that in response to a shock such as a new disruptive technology that there are a variety of zones defining the adjustment of each technology. Each zone is defined by complementary slackness and characterised by either capacity adjustment while returns are fixed, or return adjustment while capacity is fixed. As capacity of the new technology is introduced, the returns of all technologies fall, although there will be periods of price stability during which adjustment will occur through retirement of the capacity of existing technologies. Naturally, the order in which technologies adjust is dependent on the nature and utilisation of the disruptive technology being introduced and the impact it has on the full PDC.

At this point we re-introduce minimum and maximum capacity levels as discussed in Section 3.3.2. The capacity restrictions and the new definition of capacity are summarised below:

$$CAP_i - \mathrm{CAP_i^-} \geq 0 \quad \perp \ \chi_i^- \geq 0 \qquad \forall i > 0 \qquad (3.26)$$

$$\mathrm{CAP_i^+} - CAP_i \geq 0 \quad \perp \ \chi_i^+ \geq 0 \qquad \forall i > 0 \qquad (3.27)$$

$$CAP_i = \mathrm{CAP_i^0} + INV_i - RET_i \qquad \forall i > 0 \qquad (3.28)$$

The re-introduction of these constraints potentially impairs the full equilibration of the system through investment and retirement alone, and requires modification of (3.21) and (3.22) to:

$$\sum_t \mathrm{w_t} \chi_{i,t} - \mathrm{FOC_i} - \left(\chi_i^+ - \chi_i^-\right) + \chi_i^{RET} \geq 0 \quad \perp \ RET_i \geq 0 \qquad \forall i > 0 \qquad (3.29)$$

$$\mathrm{FC_i} - \sum_t \mathrm{w_t} \chi_{i,t} + \left(\chi_i^+ - \chi_i^-\right) \geq 0 \quad \perp \ INV_i \geq 0 \qquad \forall i > 0 \qquad (3.30)$$

Capacity bounds could apply impact the feasible range of both investment and retirement decisions and this is reflected above. When binding, these restrictions alter sub-period returns and optimal trade-offs,

but as before, while the trade-offs will be numerically different, there is no requirement to alter their formulation as in (3.13).

*Mothballing Capacity*

Within the context of a particular market, the balance of the mothballing/retirement decision is of significance to investors. Investors in new capacity will see returns suffer if the market and its variations are such that, with reasonable accuracy, pre-existing mothballed plant can be returned to active operation at times where prices may be generally high. Conversely, mothballing and re-instatement may allow investors to eke out additional value.

Mothballing and reinstatement provide some relief from the hitherto discussed issue of capacity inflexibility. In most standard models, the possibility of adjusting capacity in the short term is not considered. Mothballing provides owners of existing plant with an intermediate option, between maintaining full operational status and retirement. Although mothballing is commonly associated with older, perhaps inefficient plants, even technologies that are profitable and attractive to investors may be candidates for mothballing. To avoid pre-judgement, rather than being a characteristic limited to plants due for retirement, mothballing and reinstatement is an operating strategy which should be considered for all plants, including new installations.

In the limit, and given enough lead-time of the requirement for additional capacity, the operating life of a plant could predominantly be spent in a mothballed state, with sporadic reinstatement as required. Although we do not address it here, there is also the possibility that the flexibility of a particular installation is a design choice that could be optimised so that the plant concerned may be more susceptible to taking advantage of reliable variation in load conditions. An extreme example of this logic would be the case of transportable power stations that service different locations at different times of year or assist during extended breakdowns of in situ capacity. We are not suggesting this possibility as the focus of the section, however our formulation does allow for this possibility.

We consider a simplified version of mothballing in which capacity is mothballed and reinstated in specific sub-periods. While the actual nature of the decision naturally involves forecasting, and recognising the plant owner has a real option to exercise or not, we limit our investigation to the possibility of capacity being reinstated for a single sub-period, although we could adapt this to consider reinstatement by scenario, or in a more advanced model, reinstatement for multiple periods. We define the cost of mothballing capacity as $\mathrm{MC}_{\mathrm{i}}^{\mathrm{MBL}}$. Whereas retiring existing capacity saves all fixed operating costs, to maintain the possibility of reinstatement, mothballed capacity still incurs some fixed operating costs, albeit lower than when in full operation, so that $0 < \mathrm{FOC}_{\mathrm{i}}^{\mathrm{MBL}} < \mathrm{FOC}_{\mathrm{i}}$.

Before we can weigh the costs and benefits of mothballing we define both the condition under which reinstatement will occur and the profitability involved when reinstatement does occur. Reinstatement feasibility is governed by:

$$\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}} MBL_i - REI_{i,t} \geq 0 \quad \perp \quad \chi_{i,t}^{REI} \geq 0 \qquad \forall i > 0, t \qquad (3.31)$$

Here $MBL_i$ is the capacity of technology i that has been mothballed, $REI_{i,t}$ is the quantity of technology i that is reinstated for operation in sub-period t, and $0 \le a_{i,t}^{REI} \le 1$ is the proportion of mothballed capacity of technology i that is feasible to reinstate in sub-period t. When $a_{i,t}^{REI} = 0$, reinstatement is not feasible, while when $a_{i,t}^{REI} = 1$, reinstatement of all mothballed plant is feasible. Feasibility in this context reflects two central aspects of the problem. Firstly, there is the literal, or technical, feasibility of reinstatement of the plant. Secondly, there is the feasibility associated with having sufficient notice of a potential future opportunity. Without the latter, whether or not reinstatement is technically possible, it will not occur. Therefore, $a_{i,t}^{REI}$ should reflect technical issues such as a complete inability to reinstate certain technology types as well as the trade-off between the time to reinstate and the notice period associated with a potential future opportunity.

In the context of sub-periods, a high value of $a_{i,t}^{REI}$ might correspond to systems with strong predictable variations, or it might correspond to systems prone to long term variations, such as hydro based systems, where the system state changes slowly, and crises, when they occur, are prolonged and sustained with significant lead time. Conversely, where system crises are short term, such as when driven by demand spikes associated with extreme heat, there is little possibility of reinstating a plant to take advantage of such opportunities. The second aspect of feasibility could be more comprehensively addressed by a detailed modelling of the information structure of the problem. Ours is a significantly simplified approach designed to reflect a less than perfect foresight. As a result, rather than allowing all technically feasible reinstatement in a model in which foresight is not perfect, our model synthesises that approach by allowing limited reinstatement but implicitly assuming foresight is perfect. As always, where an issue is significant in a particular context, modelling resources should be reallocated towards it, and the case that this is a significant market feature, it would suggest a fuller specification of the information structure.

Given reinstatement is feasible according to (3.31), reinstatement will occur whenever the available marginal returns exceed the marginal cost of reinstatement, $MC_i^{REI}$, which is assumed constant. $MC_i^{REI}$ includes the technological costs incurred in reinstating capacity to operational status and then returning it to mothballed status, as well as an adjustment that reflects that the operating cost savings from mothballed capacity, $FOC_i - FOC_i^{MBL}$, are not available when the capacity is operating. The condition governing reinstatement is:

$$MC_i^{REI} - w_t \chi_{i,t} + \chi_{i,t}^{REI} \ge 0 \quad \perp \; REI_{i,t} \ge 0 \qquad \forall i > 0, t \tag{3.32}$$

Where the cost of reinstatement exceeds the returns available, reinstatement will not occur. Where the opposite is true, capacity will be reinstated until the marginal profitability in sub-period t falls to parity with the marginal cost of reinstatement. The exception occurs when reinstatement is restricted by the feasibility constraint (3.31) and the equilibrium is characterised by a marginal profitability that exceeds the marginal cost of reinstatement, $w_t \chi_{i,t} > MC_i^{REI}$, but is unable to be capitalised upon.

Notionally, mothballing provides the plant owner with a strip of options to reinstate the plant whenever they choose. The equilibrium marginal value of the option to reinstate technology i in period t is given by $\chi_{i,t}^{REI}$, however the validity of the equilibrium option valuation requires some investigation. If reinstatement is not constrained by (3.31) then $\chi_{i,t}^{REI} = 0$ and $\mathrm{MC}_{\mathrm{i}}^{\mathrm{REI}} - \mathrm{w}_{\mathrm{t}}\chi_{i,t} \geq 0$, so the option value is captured correctly. When reinstatement is constrained at a positive level then $\chi_{i,t}^{REI} = \mathrm{w}_{\mathrm{t}}\chi_{i,t} - \mathrm{MC}_{\mathrm{i}}^{\mathrm{REI}} \geq 0$, and again the option value is captured correctly. However, when reinstatement is constrained at zero, $\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}} = 0$ and the situation is ambiguous. Either $\mathrm{w}_{\mathrm{t}}\chi_{i,t} - \mathrm{MC}_{\mathrm{i}}^{\mathrm{REI}} > 0$ and $\chi_{i,t}^{REI} = \mathrm{w}_{\mathrm{t}}\chi_{i,t} - \mathrm{MC}_{\mathrm{i}}^{\mathrm{REI}}$, reflecting that reinstatement is desirable but not feasible, or $\mathrm{MC}_{\mathrm{i}}^{\mathrm{REI}} - \mathrm{w}_{\mathrm{t}}\chi_{i,t} \geq 0$, and $\chi_{i,t}^{REI}$ is free, reflecting reinstatement is economically undesirable. In each case the option value, $\chi_{i,t}^{REI}$, being reported is incorrect because the option to reinstate does not exist. Fortunately, this is of little import in our formulation as the weighting attached to the marginal profitability in the relevant sub-period is $\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}} = 0$, as shown in (3.33).

The profit from mothballing comprises reductions in fixed operating costs and the profitability from reinstatement. Given a one-off charge, $\mathrm{FC}^{\mathrm{MBL}}$, which we express as an annuity on the same basis as other fixed costs, and which reflects the cost of configuring a technology to accommodate reinstatement, the profitability available from configuring and optimally mothballing and reinstating a unit of capacity is:

$$\mathrm{FOC}_{\mathrm{i}} - \mathrm{FOC}_{\mathrm{i}}^{\mathrm{MBL}} + \sum_{t} \mathrm{w}_{\mathrm{t}}\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}}\chi_{i,t}^{REI} - \mathrm{FC}^{\mathrm{MBL}} \qquad \forall i > 0 \qquad (3.33)$$

The fixed operating cost savings from mothballing apply to all units of capacity, whereas the benefits from reinstatement are scaled to reflect the possibility of reinstatement being constrained. In equilibrium, in those periods where reinstatement is not constrained, $\chi_{i,t}^{REI} = 0$, and the scaling factor, $\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}}$, is not relevant. In those sub-periods in which the reinstatement is constrained, $\mathrm{a}_{\mathrm{i,t}}^{\mathrm{REI}}$ reflects the proportion of reinstatement available as a result of mothballing an additional unit of capacity and $\chi_{i,t}^{REI} \geq 0$ records the marginal profitability of reinstating capacity.

We now seek to integrate the mothballing and reinstatement conditions with the rest of the framework. For the purpose of this discussion, mothballed capacity is considered capacity, so that the definition of $CAP_i$ remains as in (3.28). We limit the extent of mothballing, $MBL_i$, with following complementarity condition:

$$CAP_i - MBL_i \geq 0 \quad \perp \chi_i^{MBL} \geq 0 \qquad \forall i > 0 \qquad (3.34)$$

Whether or not mothballing and reinstatement are profitable is not the relevant test as plant owners have other operational strategies available. In this case, the profitability of mothballing must be assessed against the profitability of normal operations. Accounting for this and the possibility of mothballing being constrained, we have the following complementarity condition governing the mothballing decision:

$$\sum_t \mathrm{w}_t \chi_{i,t} - \pi_i^{MBL} + \chi_i^{MBL} \geq 0 \quad \perp \; MBL_i \geq 0 \qquad \forall i > 0 \tag{3.35}$$

Where:

$$\pi_i^{MBL} = \mathrm{FOC_i} - \mathrm{FOC_i^{MBL}} + \sum_t \mathrm{w}_t \mathrm{a_{i,t}^{REI}} \chi_{i,t}^{REI} - \mathrm{MC^{MBL}} \qquad \forall i > 0 \tag{3.36}$$

When the marginal profitability of normal operations exceeds the marginal profitability of mothballing in equilibrium, (3.35) implies that no mothballing of that technology occurs. Where the opposite is true, and providing that the mothballing restriction in (3.34) is not binding, capacity will be mothballed until the marginal profitability of doing so equates with the marginal profitability of normal operations. If all capacity is mothballed, the marginal profitability of mothballed capacity may remain higher than the marginal profitability of normal operations. The profitability differential defines $\chi_i^{MBL} \geq 0$, to satisfy (3.35).

Mothballing and reinstatement represent the introduction of an additional operational choice for each technology. The equilibrium investment condition, as well as the equilibrium retirement condition must account for the most profitable use for each technology and not merely the spot market returns that are available. It would be convenient to express the profitability of each technology as $\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}$, where $\chi_i^{MBL}$ measures the marginal value of the option to mothball over and above the marginal profitability of spot market operation.

As with the marginal value of the reinstatement option, we can show this is the case in all but one situation, and that this situation represents a self-contained equilibrium that does not contaminate the results. From (3.35), when capacity is only partially mothballed then, from (3.34), $\chi_i^{MBL} = \pi_i^{MBL} - \sum_t \mathrm{w}_t \chi_{i,t} = 0$, reflecting the marginal option value we desire. When all built capacity is mothballed $\chi_i^{MBL} = \pi_i^{MBL} - \sum_t \mathrm{w}_t \chi_{i,t} \geq 0$, reflecting the marginal option value we desire. When mothballing and the capacity available to mothball are both zero in equilibrium, then although mothballing is constrained physically, its economic desirability is undetermined. In this case, we have $\chi_i^{MBL} = \pi_i^{MBL} - \sum_t \mathrm{w}_t \chi_{i,t} \geq 0$, reflecting the option value we desire. Where it is not $\chi_i^{MBL} = \pi_i^{MBL} - \sum_t \mathrm{w}_t \chi_{i,t} \leq 0$ and, as far as complementarity conditions (3.34) and (3.35) are concerned, the value of $\chi_i^{MBL}$ is free. In this case alone, $\chi_i^{MBL}$ does not define the option value of mothballing. However, for capacity to be zero, investment must be zero and retirements must be maximised, in which case mothballing is not an available option. Were $\chi_i^{MBL}$ to rise to a level where zero capacity of the technology concerned was no longer a feature of the equilibrium solution, mothballing would cease to be constrained at zero and $\chi_i^{MBL}$ would again define the option value of mothballing.

Accordingly, we can express the profitability of each technology as $\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}$ in the knowledge that the only scenario in which this is not strictly the case, is a self-contained solution. We amend (3.29) and (3.30), which are respectively the equilibrium retirement condition and the

equilibrium investment condition, as follows to reflect the potential of mothballing and reinstatement to adjust earnings:

$$\left(\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}\right) - \mathrm{FOC}_i - \left(\chi_i^+ - \chi_i^-\right) + \chi_i^{RET} \geq 0 \quad \perp \quad RET_i \geq 0 \qquad \forall i > 0, t \qquad (3.37)$$

$$\mathrm{FC}_i - \left(\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}\right) + \left(\chi_i^+ - \chi_i^-\right) \geq 0 \quad \perp \quad INV_i \geq 0 \qquad \forall i > 0, t \qquad (3.38)$$

The logic supporting these conditions remains unaltered although the definition of profit in each case has been altered. From (3.37), the effect of (weakly) greater profits is to (weakly) reduce the amount of capacity retired, while from (3.38) investment will be (weakly) higher than in a model in which mothballing is not considered. Unsurprisingly, the net result of the identification of an additional operational option, in this case mothballing and reinstatement, is weakly greater investment supported by the potential of further profit that encourages capacity to be retained or invested in for that purpose. Overall, the level of actual capacity installed is ambiguous because of the potential mothballing of pre-existing, economically justifiable, plants. When viewed at an individual technological level, the ability of competing technologies to do the same, or exhibit even greater flexibility, may counteract this effect for an individual technology.

Turning to the spot market implications of mothballing and reinstatement, the level of capacity available for generation is dependent on the level of mothballing and reinstatements within a particular sub-period, so that condition (3.29) becomes:

$$CAP_{i,t} - MBL_i + REI_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi_{i,r,t}^+ \geq 0 \qquad \forall i > 0, r, t \qquad (3.39)$$

This change permits reinstated capacity to augment the underlying capacity and be used for generation. As before no change is required in the definition of sub-period optimal trade-offs.

Ours has been a simplified evaluation of mothballing and retirement. Mothballing provides an additional operating mode for each technology. Once a plant is mothballed, the owner of the plant has, in effect, a strip of options to reinstate their capacity whenever it is profitable and feasible to do so. The cumulative value of these opportunities defines the marginal profitability of mothballing, and by comparison with the profitability of spot market operations, the marginal value of mothballing as an option. Most importantly, while the optimal trade-offs that define the equilibrium will adjust, that adjustment is accommodated in the definition of the sub-period profitability, so no further adjustment to the optimal trade-off condition is required.

The evaluation of mothballing decisions rests significantly on the evaluation of future profitability, which is not certain. Whether a future circumstance is profitable depends on the suitability of mothballed plants to respond in a practical, as well as economic, sense. To be captured, future opportunities must be reliably forecasted with a lead-time sufficient for the operator to mobilise the plant, and of sufficient duration for the generator to recover any fixed costs of reinstatement. A detailed examination of operator forecasting would require a representation of the evolution of uncertainty in the system, perhaps through a Markov chain or decision process, and recognition of the nature of the decision as being that of a real option. This requirement stems from the need for

operational decisions to be made about mothballing and reinstating plant, which in turn are based on the distribution of future system states, the lead time to feasibly complete those actions and the length of time the system will remain in that desirable/undesirable state.

Finally, an even more sophisticated view of mothballing and/or retirement might also attribute other value to an existing plant. For example, it may provide risk reduction, or deterrence value in a future state. We have not addressed this but it is certainly an issue of interest to any investors who foresee a significant portion of equilibrium cost recovery occurring in times of system stress, and who would not want to see that disrupted by the reinstatement of older plants that might prevent pricing from reaching shortage pricing levels. This is particularly relevant in systems where load is falling, creating an overhang of capacity.

## *3.4 Energy Limits & Storage*

### 3.4.1 Introduction

Energy limits may apply at the plant level, such as for hydro-generation, where individual hydro plants may have their own energy limits, or at a technological level, or even across a range of technologies that share the same fuel. To accommodate this range of possibilities, unique technological opportunities could be formulated as distinct technologies, and/or joint constraints could be applied to those technologies that share the same fuel. We focus on the intermediate case, and assume each technology has a distinct fuel, subject to its own energy limit.

Depending on the reason for the limit, energy limits can be deterministic, arising from contractual arrangements such as those defined by take or pay fuel contracts, or stochastic, arising from the interaction between nature and generation technologies, such as water inflows into hydro-electric generation schemes. Whatever the source of limitation, energy limits create an opportunity cost for the limited energy source or fuel. Depending on one's perspective, fuel should be used when electricity prices are highest to reap maximum profit, or equivalently fuel should be used to reduce usage of more expensive alternatives. Although fuel allocation decisions are more complex in the stochastic case, the allocation of fuel depends on the capacity of the technology using the fuel and the ability of firms to store fuel for use in those periods that in which it is most valuable, whether energy arrivals are stochastic or deterministic.

Figure 24 makes clear that if fuel was available in whatever quantity was desired, a comparable technology without energy limits would operate over a wider utilisation range, supply a greater portion of total energy, and have a greater installed capacity in the optimal plant mix relative to an equivalent energy limited technology. Calculation of optimal energy use requires a well-defined LDC, and the interaction between the energy limit and the LDC implies an equilibrium fuel value, $\varepsilon_i$. In a deterministic setting, $\varepsilon_i$ adjusts until the available fuel is exactly used, at which point the capacity requirement is defined as shown in Figure 24. The optimal value of such fuel is set so that consumption will exactly equal the amount available. If the opportunity cost of fuel were assessed at too high a level, fuel would be left over, whereas if it were set too low, fuel would run out having not necessarily been used in an optimal fashion.

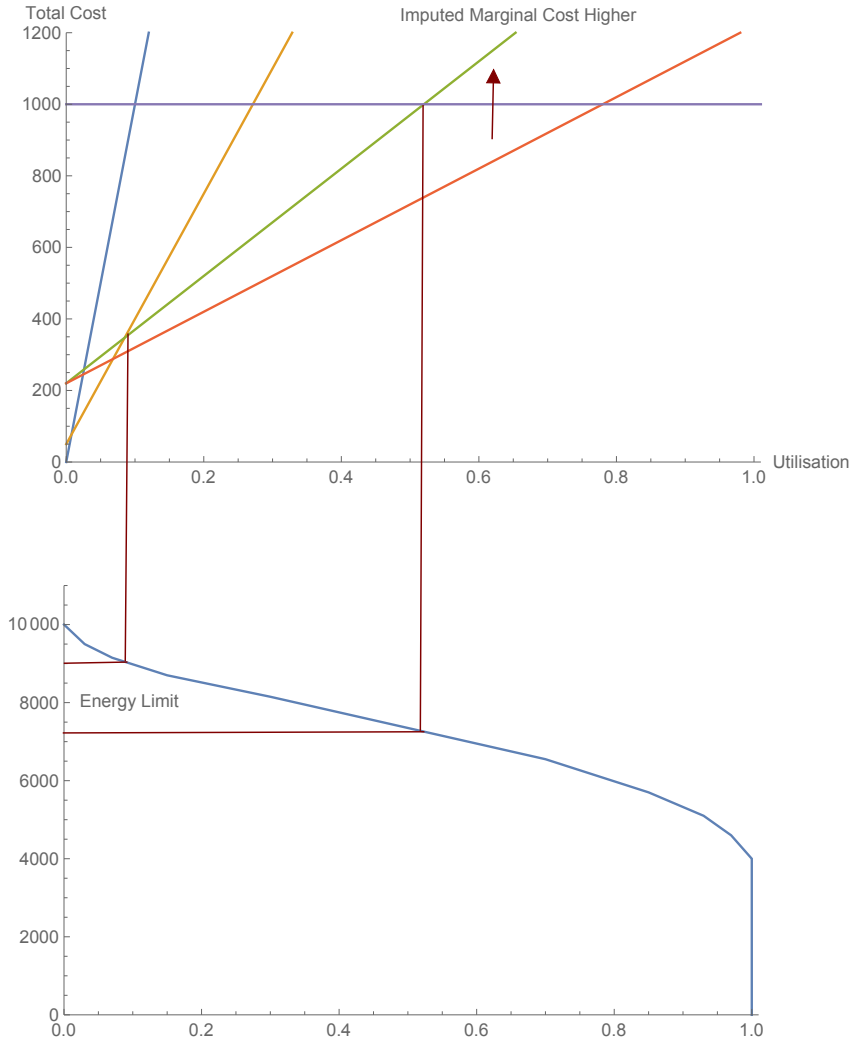



**Figure 24: Energy Limited Technologies**

The value of installed capacity, and the return it can generate, is dependent on the amount of energy that can be produced, which is limited not only by the rated capacity of the plant but the amount of fuel that is available for the period under consideration (Bernard & Chatel, 1984). Slightly generalising that result, if a technology were available in any quantity, but had a known energy limit per MW of capacity, the optimal energy/capacity ratio (or plant output factor) for that technology could be determined by re-valuing the fuel resource until the optimal plant mix would use the fuel available, and no more.

In terms of a single technology, i, that has a marginal operating range in the interior of the linear LDC section defined by $\{u_k, u_{k+1}\}$, the above adjustment process may be interpreted as the solution to the following set of simultaneous equations:

$$E_i = \frac{1}{2}\left(\frac{L_k - L_{k+1}}{u_{k+1} - u_k}\right)\left(\left(u^e_{n+1}\right)^2 - \left(u^e_n\right)^2\right) \tag{3.40}$$

$$u^{e}_{n+1} = \frac{\mathrm{FC_{i+}} - \mathrm{FC_{i}}}{\varepsilon_i - \mathrm{MC_{i+}}} \tag{3.41}$$

$$u^{e}_{n} = \frac{\mathrm{FC_{i}} - \mathrm{FC_{i-}}}{\mathrm{MC_{i-}} - \varepsilon_i} \tag{3.42}$$

Once again, i+, and i- refer to the neighbouring technologies as shown, which are not necessarily the technologies i+1, or i-1, as indexed in the original problem. Equation (3.40) defines the relationship between the energy limit and the marginal utilisation range of technology i, in terms of the (negative) slope of the LDC. We take this opportunity to reiterate the form of (3.40) reflects a linear, or piecewise linear, LDC and enables the calculation of the energy content of the LDC without compromising the definition of the capacity requirement as is necessary when modelling a piecewise constant LDC. This defines the width of the marginal utilisation range which, in conjunction with (3.41) and (3.42) implies $\varepsilon_i$, the opportunity cost of fuel.

To get both the optimal capacity/energy ratio and the optimal capacity investment when energy is limited and capacity is flexible, we must solve for the opportunity cost of the fuel and the development option simultaneously. As shown in Figure 25, when capacity is fixed and energy is limited, then both the opportunity cost of fuel, $\varepsilon_i$, and the opportunity cost of existing capacity, $\chi_i$, will adjust.

In terms of a single technology, i, whose marginal operating range lies in the interior of the range $\{\mathrm{u_k}, \mathrm{u_{k+1}}\}$, then in a particular period, that adjustment process may be interpreted as the solution to the following simultaneous equations:

$$CAP_i = \left( \frac{\mathrm{L_k} - \mathrm{L_{k+1}}}{\mathrm{u_{k+1}} - \mathrm{u_k}} \right) \left( u^{e}_{n+1} - u^{e}_{n} \right) \tag{3.43}$$

$$\mathrm{E_i} = \frac{1}{2} \left( \frac{\mathrm{L_k} - \mathrm{L_{k+1}}}{\mathrm{u_{k+1}} - \mathrm{u_k}} \right) \left( \left( u^{e}_{n+1} \right)^2 - \left( u^{e}_{n} \right)^2 \right) \tag{3.44}$$

$$u^{e}_{n+1} = \frac{\mathrm{FC_{i+}} - \chi_i}{\varepsilon_i - \mathrm{MC_{i+}}} \tag{3.45}$$

$$u^{e}_{n} = \frac{\chi_i - \mathrm{FC_{i-}}}{\mathrm{MC_{i-1}} - \varepsilon_i} \tag{3.46}$$

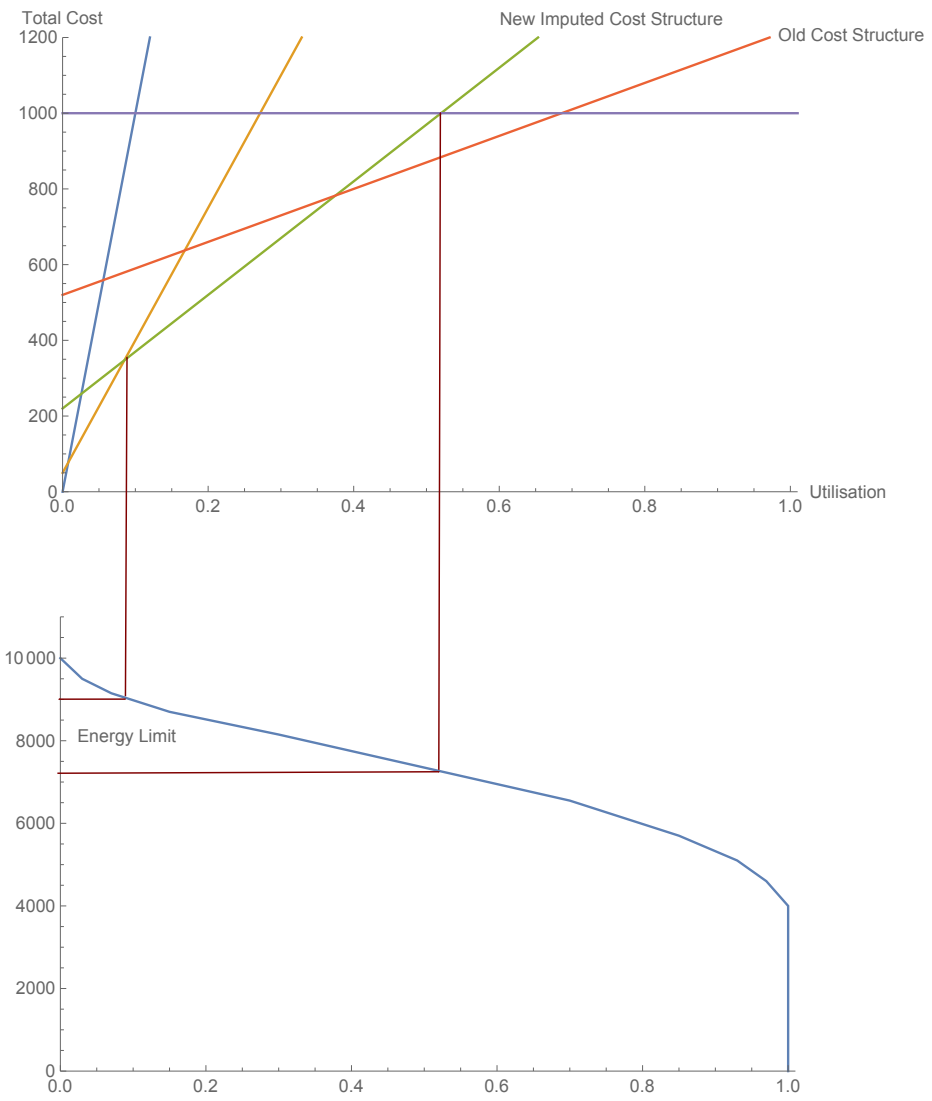



**Figure 25: Energy & Capacity Limited Technologies**

The boundaries of the marginal utilisation range, in which it is more cost effective to use technology i rather than neighbouring technologies, are defined by (3.45) and (3.46).

Technologies with the same capital:energy ratio will have the same marginal fuel value, provided that:

- Generators are risk neutral. This is a sufficient condition. A necessary condition would be that there is no difference in marginal fuel values between periods due to risk.
- Storage is not constrained. If storage is constrained then the energy limit for the entire period will not necessarily be able to be allocated optimally across all sub-periods. The result of this is diverging marginal fuel values between sub-periods.

These results are unsurprising, as for two technologies sharing a capacity:energy ratio to precisely exhaust a limited fuel supply, they must have the same average utilisation level. We can demonstrate this by considering the capacity:energy ratio from (3.43) and (3.44), which simplifies to:

$$\frac{CAP_i}{\mathrm{E_i}} = \frac{u^e_{n+1} + u^e_n}{2} = Avg\,Utilisation \tag{3.47}$$

As two technologies with the same average utilisation must have overlapping utilisation ranges, it is clear from screening curve analysis that they must also share a marginal fuel value.

### 3.4.2 Deterministic Energy Limits

*Annual Fuel Arrival*

We begin by considering fuel that arrives annually in a deterministic quantity, albeit possibly chosen by the plant operator. We assume that fuel and energy are related by a constant scaling factor, reflecting a constant conversion rate between each. This enables fuel limits to be considered as energy limits. To address conversion rates, another set of equations would be required in the model to provide translation between fuel limits and energy limits. Given a maximum available fuel quantity of $\mathrm{E_i^+}$, the feasibility of an energy use pattern, consisting of distributing $E_{i,t} \geq 0$ to each sub-period t for use by technology i, is governed by the simple constraint:

$$\mathrm{E_i^+} - \sum_t E_{i,t} \geq 0 \quad \perp \quad \varepsilon_i \geq 0 \qquad \forall i \tag{3.48}$$

Implicit in the single arrival assumption is that sufficient storage exists, and therefore no restrictions on transfer are applicable. We also assume that there is free disposal so that whenever fuel is available in abundance, the opportunity cost of fuel has a floor, $\varepsilon_i = 0$. At other times, when fuel availability is limited below freely desired levels, $\varepsilon_i > 0$ which reflects the opportunity cost of that fuel in terms of its energy value.

The energy use definition (3.40) applies within an LDC section or when the LDC is linear and therefore has a constant slope. That range spans only one set of endogenous utilisation levels, but our framework incorporates a more general LDC which is piecewise linear, requiring an updated definition of energy use, that correctly identifies energy use over operating ranges potentially spanning several slices of the LDC. We define each LDC slice by its marginal utilisation range $\{u_r, u_{r+1}\}$. For each slice of the LDC, the total energy use for the slice is:

$$E_{r,t} = \frac{1}{2}\left(L_{r,t} - L_{r+1,t}\right)\left(u_{r+1,t} + u_{r,t}\right) \qquad \forall r,t \tag{3.49}$$

The energy use for each technology in each slice of the LDC can be defined directly using generation variables, $GEN_{i,r,t}$, instead of load variables:

$$E_{i,r,t} = \frac{1}{2}\left(GEN_{i,r,t} - GEN_{i,r+1,t}\right)\left(u_{r+1,t} + u_{r,t}\right) \qquad \forall i > 0, r, t \tag{3.50}$$

The substitution of load, which is a parameter, with generation, which is a decision variable, introduces a second degree of freedom into this equation. The energy limit should be satisfied by adjustment in the opportunity cost of energy. However, this change appears to enable the energy constraint to be

satisfied by changes in generation, rather than the intended adjustment of the energy price. Fortunately, the framework is robust to this possibility. The inclusion of sub-periods in the model re-orients the definition of marginal trade-offs so that when the total cost of supplying an increment of load is equal between two technologies, as it is at an optimal trade-off such as $u_{r,t}$, potential ambiguity is resolved by preventing higher marginal cost technologies from generating at $u_{r,t}$. In combination with the standard market clearing constraint, this implies that generation is not free at the optimal trade-off $u_{r,t}$.

Summing over LDC slices and restating as an equilibrium complementarity condition we have:

$$E_{i,t} - \frac{1}{2}\sum_{r=0}^{R-1}\left[\left(GEN_{i,r,t} - GEN_{i,r+1,t}\right)\left(u_{r+1,t} + u_{r,t}\right)\right] \geq 0 \quad \perp \quad \varepsilon_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.51)$$

While it is possible to substitute (3.51) into (3.48) we prefer to leave the intermediate variable in the model.

We have defined $\varepsilon_i$ as the opportunity cost of fuel which values fuel at whatever level is necessary to use all available fuel, but we must define the maximum energy limit, $\mathrm{E}_{\mathrm{i}}^{+}$, in such a way that it comports with potential rationales for an energy limit, such as take or pay contracts. Fuels that arrive naturally such as hydro inflows have $\mathrm{MC}_{\mathrm{i}} = 0$, but others with $\mathrm{MC}_{\mathrm{i}} > 0$ may also be limited. For such fuels, it is possible that the opportunity cost $\varepsilon_i < \mathrm{MC}_{\mathrm{i}}$, so that where fuel has $\mathrm{MC}_{\mathrm{i}} > 0$ the energy limit is contingent on the opportunity cost of fuel, $\varepsilon_i$, being at or above the fuel price.

We assume that the energy limit, $E_i$, is variable in a single step between zero and the maximum energy limit, $\mathrm{E}_{\mathrm{i}}^{+}$. This could be generalised to include further tranches where various fuel contract options, or extensions are available. The following set of complementarity conditions reflect the nature of a simple take or pay contract, and in doing so mirror the basic structure of market clearing conditions, as they would be with a single technology:

$$\mathrm{MC}_{\mathrm{i}} - \varepsilon_{i,t} + \eta_i^f \geq 0 \quad \perp \quad E_i \geq 0 \qquad \forall i > 0, t \qquad (3.52)$$

$$\mathrm{E}_{\mathrm{i}}^{+} - E_i \geq 0 \quad \perp \quad \eta_i^f \geq 0 \qquad \forall i > 0 \qquad (3.53)$$

Where $\varepsilon_{i,t} < \mathrm{MC}_{\mathrm{i}}$, then energy availability, $E_i = 0$, as no fuel will be purchased at a cost greater than its value. Where $\varepsilon_{i,t} > \mathrm{MC}_{\mathrm{i}}$, notably including cases where the fuel has limited availability but is available at no cost, then $E_i = \mathrm{E}_{\mathrm{i}}^{+}$ as $\eta_i^f = \varepsilon_{i,t} - MC_{i,t} > 0$, implying fuel purchases or arrivals are weakly profitable. Where $\varepsilon_{i,t} = \mathrm{MC}_{\mathrm{i}}$, the firm can adjust fuel purchases freely up to the maximum energy limit $\mathrm{E}_{\mathrm{i}}^{+}$, after which point the adjustment reverts to the opportunity cost of fuel so that $\varepsilon_{i,t} > \mathrm{MC}_{\mathrm{i}}$.

The optimal allocation of energy between sub-periods requires that the global opportunity cost of fuel, representing the most valuable use of fuel, is at least the opportunity cost of fuel in each sub-period:

$$\varepsilon_i - \varepsilon_{i,t} \geq 0 \quad \perp \quad E_{i,t} \geq 0 \qquad \forall i > 0 \qquad (3.54)$$

When $\varepsilon_i > \varepsilon_{i,t}$, the opportunity cost of fuel to the system is greater than in sub-period t, implying that there should not be an allocation of limited fuel for use in that sub-period. If there were, fuel could profitably be reallocated from sub-period t to whichever sub-period provides support for the global opportunity cost of fuel. Accordingly, where there is an allocation in a sub-period, it must be the case that the opportunity cost of fuel in that sub-period equates with the global opportunity cost, $\varepsilon_i = \varepsilon_{i,t}$.

From the perspective of the market clearance process, the consideration of energy limits renders the marginal cost of fuel an intermediate variable, internal to the firm concerned. The fuel user values the fuel according to its opportunity cost. We recognise this in a perfectly competitive market clearance by substituting marginal cost, $\mathrm{MC_i}$, for the opportunity cost of fuel, $\varepsilon_{i,t}$, as it is this price that sets the market price whenever the technology concerned is marginal.

$$-\lambda_{r,t} + \varepsilon_{i,t} + \varphi^{+}_{i,r,t}\Big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (3.55)$$

The optimal trade-off conditions must also be amended to reflect the opportunity costs of fuel:

$$\chi_{i,t} - \chi_{j,t} - (\varepsilon_{j,t} - \varepsilon_{i,t})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \qquad (3.56)$$

As shown in (3.57) and (3.58), the equilibrium investment condition remains the same. There is no need to adjust the formulation with respect to the definition of $\chi_{i,t}$, as the implications of energy limits in each sub-period are recorded through changes in profitability arising from adjusted market prices and utilisation ranges.

$$\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} + \left(\chi^{+}_i - \chi^{-}_i\right) \geq 0 \quad \perp \quad INV_i \geq 0 \qquad \forall i > 0 \qquad (3.57)$$

$$\sum_t \mathrm{w_t}\chi_{i,t} - \mathrm{FOC_i} - \left(\chi^{+}_i - \chi^{-}_i\right) \geq 0 \quad \perp \quad RET_i \geq 0 \qquad \forall i > 0 \qquad (3.58)$$

While the investment and retirement conditions (3.57) and (3.58) remain the same, we note that we are implicitly assuming that the capacity choice for each generation technology, along with its storage or means of acquiring the limited fuel, are able to be treated separately. In such a case, the value of an additional unit of storage is measured by the weighted difference between the opportunity and marginal cost of the fuel. In equilibrium, barring the imposition of any limits, this should be equal to the cost of an incremental unit of storage. In some cases, such as hydroelectric generation capacity, capacity and storage capacity may be entwined. This implies constraints relating capacity and storage choices and results in the introduction of additional variables into the optimal investment condition for each. Where capacity and storage capacity must move in lock step, we could modify the investment

condition to reflect the combined cost of installation and the additional fuel value created by the existence of storage.

The equilibration process exhibits a series of capacity and marginal capacity value adjustments, taken in succession, reflecting the mutually exclusive zones in which capacity and marginal capacity values are flexible. It is helpful to view the process of equilibration as a function of the opportunity cost of fuel. When an energy limit is applied, the adjustment is a function of the opportunity cost of fuel, $\varepsilon_{i,t}$, which adjusts until the limited fuel is exactly used. Ceteris paribus, a lower/higher maximum energy limit, $\mathrm{E}_i^+$, implies a higher/lower $\varepsilon_i$, and either a lower/higher equilibrium average marginal capacity value, $\sum_t \mathrm{w_t}\chi_{i,t}$, or a lower/higher capacity, $CAP_i$. Where $INV_i > 0$ in the absence of an energy limit, lower energy limits results in less equilibrium investment and therefore total capacity until either $CAP_i = \mathrm{CAP_i^-}$ or $INV_i = 0$, leaving only pre-existing capacity. In the first case, $CAP_i = \mathrm{CAP_i^-}$, and capacity can no longer adjust downwards, resulting in a differential between fixed costs and marginal capacity values, as explained in Section 3.3.2. In the second case, in which $INV_i = 0$, only existing capacity remains. The process of marginal capacity value adjustment continues until eventually the marginal capacity value will reach the level of fixed operating costs, $\mathrm{FOC_i}$, and capacity retirement will be incentivised. When capacity retirements occur, the marginal value of capacity remains constant, as the required adjustment to accommodate the energy limit is offset by adjustments in the capital stock. Eventually, if a strict enough energy limit is applied, all capacity will be retired and a differential, between $\mathrm{FOC_i}$ and $\sum_t \mathrm{w_t}\chi_{i,t}$, will remain. That differential will define the marginal subsidy that would be required to prevent the retirement of the last unit of capacity of technology i.

Many technologies may be limited, and the energy limits of one technology will interact with those of others. We could imagine the case of a thermal plant with a contractually limited fuel supply looking to take advantage of lower than usual hydro inflows, or a hydro generator in one region looking to take advantage of low inflows in another region. Because the opportunity cost of one fuel is pegged to the next best alternative use, changing the energy limit for one technology implies a re-assessment of the value of other limited resources and simultaneous determination of the opportunity costs of all limited fuels. It follows from (3.56) that this dynamic will impact optimal trade-offs and investment, with the overall effect on the optimal utilisation of each technology dependent on which opportunity cost effects are stronger.

*Sub-Period Fuel Arrival*

We now consider the arrival of energy in each sub-period. This could reflect the delivery structure embedded in take or pay arrangements, or natural considerations such as seasonal inflow patterns, for example. It is implicit that single fuel deliveries can be stored. Our focus here is demonstrating the impact of storage on the structure of fuel allocation, and how that can be incorporated into our

structure. To illustrate the nature of the decision we assume future fuel availability is known with certainty. In Appendix 7.3 we briefly describe how stochastic fuel availability could be considered.

Irrespective of whether inflows are limited by a natural process or contractual terms, we define a maximum inflow of fuel for technology i in period t of $\mathrm{INF}^{+}_{\mathrm{i,t}}$. As before, we generalise the definition of energy inflows to reflect the reality of non-zero fuel costs, such as in a take or pay contract, this time recognising the potential for different, perhaps seasonal, fuel costs in each sub-period:

$$\mathrm{MC}_{\mathrm{i,t}} - \varepsilon_{i,t} + \eta_{i,t} \geq 0 \quad \perp \quad REL_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.59)$$

$$\mathrm{INF}^{+}_{\mathrm{i,t}} - INF_{i,t} \geq 0 \quad \perp \quad \eta_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.60)$$

The nature of (3.60) depends on the precise situation being considered. As written this constraint envisages a fuel supply contract with the purchaser having the option to purchase fuel in a quantity up to the maximum specified per period at a scheduled cost. Certain contracts may allow a certain degree of swing between periods and they would require modification of the constraint. However, in the case described, where $\varepsilon_{i,t} < \mathrm{MC}_{\mathrm{i,t}}$, energy availability, $INF_{i,t} = 0$, as no fuel will be purchased at a cost greater than its value. Where $\varepsilon_{i,t} > \mathrm{MC}_{\mathrm{i,t}}$, including when fuel inflows are limited but free, available energy, $INF_{i,t} = \mathrm{INF}^{+}_{\mathrm{i,t}}$ as $\eta_{i,t} = \varepsilon_{i,t} - MC_{i,t} > 0$, implying fuel purchases are profitable. Where $\varepsilon_{i,t} = \mathrm{MC}_{\mathrm{i,t}}$, the firm can adjust fuel purchases freely up to the maximum energy limit $\mathrm{INF}^{+}_{\mathrm{i,t}}$, after which point the opportunity cost of fuel $\varepsilon_{i,t} > \mathrm{MC}_{\mathrm{i,t}}$.

The release in sub-period t must obey (3.61), the left hand condition of which will hold with equality whenever fuel has value, and will permit spillage or wastage only when fuel has no value.

$$REL_{i,t} - \frac{1}{2}\sum_{r=0}^{R-1}\left[\left(GEN_{i,r,t} - GEN_{i,r+1,t}\right)\left(u_{r+1,t} + u_{r,t}\right)\right] \geq 0 \quad \perp \quad \varepsilon_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.61)$$

The ability to store multiple periods of fuel arrivals is no longer implicitly guaranteed, and therefore neither is the ability to freely transfer fuel between sub-periods. Instead of a global constraint associated with a single delivery of a limited energy source, the following inequality defines the fuel storage at the end of sub-period t, $STOR_{i,t}$, in terms of energy inflows, $INF_{i,t}$, energy outflows, $REL_{i,t}$, and opening storage, $STOR_{i,0}$.

$$STOR_{i,0} + INF_{i,t} - REL_{i,t} = STOR_{i,t} \qquad \forall i > 0, t \qquad (3.62)$$

We note that until the introduction of inter-temporal constraints like (3.62), other sub-period decompositions, such as by time of day, or weekday and weekend, are also possible. The introduction of inter-temporal constraints such as (3.62) implicitly requires the sub-periods chosen to be contiguous to ensure intertemporal restrictions are meaningful.

With an initial storage level at the start of sub-period 1, $\mathrm{STOR}_{\mathrm{i,0}}$, the storage balance equation (3.62) can be expressed as a complementarity condition:

$$\mathrm{STOR}_{i,0}\Big|_{t=1} + STOR_{i,t-1}\Big|_{t>1} + INF_{i,t} - REL_{i,t} - STOR_{i,t} = 0 \quad \perp \quad \gamma_{i,t} \; free \qquad \forall i > 0, t \qquad (3.63)$$

Here $\gamma_{i,t}$ represents the marginal value of stored fuel at the end of sub-period t, or equivalently at the beginning of sub-period t+1. At the beginning of each sub-period, the operator chooses the quantity of fuel, $REL_{i,t}$ , to consume during the sub-period. This amounts to a choice between using fuel in the current period and conserving it for future use. In the absence of discounting and with unlimited storage, the marginal value of stored fuel will equalise across all sub-periods. The marginal value of release will typically equate to the marginal value of stored fuel although will whenever there is no release in a sub-period, a disparity between the marginal value of release and the marginal value of storage will be observed. Therefore, we can no longer claim that the marginal value of release should equalise across all sub-periods. The following complementarity condition describes the various scenarios:

$$\gamma_{i,t} - \varepsilon_{i,t} \geq 0 \quad \perp \quad REL_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.64)$$

Where $\gamma_{i,t} > \varepsilon_{i,t}$ , the release of fuel for generation in the current period, $REL_{i,t} = 0$ in equilibrium, as the marginal value of stored fuel at the end of the sub-period exceeds the marginal value of releasing fuel during the current period. When fuel is released, then $\gamma_{i,t} = \varepsilon_{i,t}$ , so the marginal value of storing fuel equates with the marginal value of release in the current period. If $\gamma_{i,t} < \varepsilon_{i,t}$ , then the marginal value of stored fuel is lower than the marginal value of release, incentivising more fuel to be released in the current period rather than saved for a less valuable future use.

We specify storage at the technological level and define maximum and minimum fuel storage as $\mathrm{STOR}^{+}_{i,t}$ and $\mathrm{STOR}^{-}_{i,t}$ respectively. Where storage limits are constant across sub-periods we could define storage in net terms, deducting that lower bound. But as seasonal implications or storage buffer zones at either the top or the bottom of the storage system might be significant, we adopt a more general approach that allows for flexible storage limits:

$$\mathrm{STOR}^{-}_{i,t} \leq STOR_{i,t} \leq \mathrm{STOR}^{+}_{i,t} \qquad \forall i > 0, t \qquad (3.65)$$

In complementarity terms, we have:

$$STOR_{i,t} - \mathrm{STOR}^{-}_{i,t} \geq 0 \quad \perp \quad \gamma^{-}_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.66)$$

$$\mathrm{STOR}^{+}_{i,t} - STOR_{i,t} \geq 0 \quad \perp \quad \gamma^{+}_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (3.67)$$

We can describe the inter-temporal linkages between optimal marginal stored fuel values with the following condition:

$$\gamma_{i,t} - \gamma_{i,t+1} + \gamma^{+}_{i,t} - \gamma^{-}_{i,t} = 0 \quad \perp \quad STOR_{i,t} \; free \qquad \forall i > 0, t \qquad (3.68)$$

Where storage limits are not binding we have $\gamma_{i,t} = \gamma_{i,t+1}$ , so that (3.68) describes the equalisation of marginal stored fuel values across contiguous sub-periods that are unaffected by storage limits. Where

$\gamma_{i,t} > \gamma_{i,t+1}$ in equilibrium, the marginal value of fuel stored at the end of period t exceeds the value of fuel stored at the end of period t+1. That state can only represent equilibrium when storage is at its minimum bound, with $\gamma^{-}_{i,t} = \gamma_{i,t} - \gamma_{i,t+1} > 0$. Conversely, where $\gamma_{i,t} < \gamma_{i,t+1}$ in equilibrium, the marginal value of stored fuel at the end of sub-period t is lower than the marginal value of stored fuel at the end of sub-period t+1. That state can only represent equilibrium when storage is at its maximum bound, with $\gamma^{+}_{i,t} = \gamma_{i,t+1} - \gamma_{i,t} > 0$. Binding storage limits prevent the unfettered transfer of energy between periods, and impair the optimal redistribution of fuel between sub-periods. In the limit, the absence of storage affords the generator no ability to maximise the value of the fuel source, and requires them to use fuel deliveries or inflows entirely within the sub-period they arrive, or waste them.

Finally, without an obligation to preserve the opening resource in some way, the resource will not be consumed in a fashion that is sustainable, which it must be in equilibrium if the solution is to be consistent. There are several ways to require the resource to be used sustainably. Perhaps the most basic approach is to restrict the end storage to the same value as beginning storage. A more adaptable approach is to generalise the storage requirement and use a value function to ascribe value to the resource at the end of the period. Without ascribing any particular motivation for doing so, we define $V_i\left(STOR_{i,T}\right)$ as that value function and $V_i^{/}\left(STOR_{i,T}\right)$ as the marginal value function associated with it. The complementarity condition (3.68) governing the progression of fuel values becomes:

$$\gamma_{i,t} - \gamma_{i,t+1}\Big|_{t<T} - V_i^{/}\left(STOR_{i,T}\right)\Big|_{t=T} + \gamma^{+}_{i,t} - \gamma^{-}_{i,t} = 0 \quad \perp \quad STOR_{i,t}\ free \qquad \forall i > 0, t \tag{3.69}$$

The market clearing conditions define the marginal value of release in conjunction with market clearing conditions and technological cost structures, but the relativity of sub-period marginal stored fuel values is disciplined by (3.69), with the end of period value function ultimately anchoring the overall value. In practice this value function may also be based on optimal resource usage, and therefore be endogenous. Further iterations of the model could be conducted to ensure the end of period fuel value function is consistent with the implied start of period fuel value.

## *3.5 Configurable Technologies*

### 3.5.1 Introduction

Typical competitive screening curve analysis suggests that certain technologies will be marginal across a wide range of utilisation levels. At the time of investment, firms have the opportunity to optimise the specific technology they are purchasing. Later in the lifespan of the plant, they may make maintenance choices that have implications for generation efficiency. In terms of the screening curve diagram, this means that there is no longer a single, well defined, "capital cost" for each technology, but a range of capital costs corresponding to more or less efficient variations of that technology. In principle, this introduces another dimension into the analysis, in which more expensive technological configurations with higher capital costs per installed MW are more efficient and are rewarded with lower marginal costs.

As with any range of technologies an investor, or a market full of investors, will potentially select a combination of technologies to service a particular load distribution. This is also true when considering individual variations of a single underlying technology. All variations that are not dominated across the relevant utilisation range will be selected as part of the optimal plant mix. To be clear, we are not considering the optimisation of a single investment, in which we choose, perhaps for reasons of decision discreteness, the configuration we most desire. Instead, we are considering an equilibrium in which investors select from a range of possible technologies. As with screening curve analysis all non-dominated technologies will have a place in the equilibrium plant mix, and in this case, this means all non-dominated configurations will have a place in the equilibrium plant mix. Accordingly, investors will observe a more nuanced PDC.

Our approach calls for the integration of non-linear and linear cost structures and effects the definition and number of optimal trade-offs and critical utilisation levels. In addition to the standard piecewise constant structure of the PDC, the PDC features linear segments which reflect the cost for a continuum of different configurations. This in turn results in a more detailed definition of returns for conventional technologies and each of the individual incarnations of each configurable technology. As a result, the modelling process is rather arduous, and we do not expect what follows to be implemented. However, the development provides several insights for market structure as well as showing a rarely seen extension of the capability of complementarity constraints.

### 3.5.2 Defining a Class of Configurable Technologies

For some technologies, the configuration options may be limited and discrete. In these cases the various configurations can be viewed as separate technologies in our analysis. But for other technologies, or where technological progress will occur, the full range of configuration permutations may be large or uncertain and investors seeking to assess the equilibrium role of a broad technological category will almost certainly not be in a position to specify all options available at the time of investment, let alone those available in the relevant future. We therefore introduce a pseudo-technology that summarises the spectrum of available options with an optimised total cost curve. This implicitly assumes investment in infinitesimal increments of efficiency are possible, so that the configuration options can be summarised by a continuous relationship between higher capital costs and lower marginal costs. Capacity of the pseudo-technology does not have the same interpretation as standard capacity in this framework. Rather than quantifying the extent of installation of a particular homogeneous capacity, the capacity of the pseudo-technology defines the total installation across an optimised and feasible range of configurations.

To simplify exposition, we address this issue at a global level, rather than a sub-period or scenario level. The approach taken highlights some of the inherent logic of configuration choices as well as further use of complementarity conditions to enforce logical conditions. We begin by partitioning available technologies into configurable and non-configurable technologies. Each configurable technology is specified defined by two limiting configurations, the most efficient configuration defined by $(FC^{+}, MC^{-})$ and the least efficient configuration, defined by $(FC^{-}, MC^{+})$. In turn, we further partition each configurable technology into two new technologies; one with a non-

linear cost structure to represent a pseudo cost function in regions where configuration is possible, and the other, a linear cost structure, to representing the limiting, and most capital intensive configuration of that technology. The limiting technology is parameterised by ($\mathrm{FC}^{+}, \mathrm{MC}^{-}$). Technologies with linear cost structures, whether they are the limiting case of a configurable technology, or non-configurable, are identified by $a_i^{cfg}=0$, whereas non-linear cost structures are identified with $a_i^{cfg}=1$. Ultimately all technologies are formulated as a single set so non-configurable technologies are presented as a special case of quadratic technologies with $\mathrm{MC}_i = \mathrm{MC}_i^{+} = \mathrm{MC}_i^{-}$, and $\mathrm{FC}_i = \mathrm{FC}_i^{+} = \mathrm{FC}_i^{-}$.

For each configurable technology, we define the nature of the trade-off between higher investment and higher efficiency. To do so we define a configuration variable, $CFG_i$, that defines the degree of configuration in the range $0 \le CFG_i \le 1$. The relative adjustment of fixed and marginal costs is determined by the actual economies, or diseconomies, of investment in production efficiency. For the purpose of our exposition, we assume the following adjustment rates:

$$MC_i = \mathrm{MC}_i^{+} - CFG_i\left(\mathrm{MC}_i^{+} - \mathrm{MC}_i^{-}\right) \qquad \forall i > 0 \qquad (3.70)$$

$$FC_i = \mathrm{FC}_i^{-} + CFG_i^{2}\left(\mathrm{FC}_i^{+} - \mathrm{FC}_i^{-}\right) \qquad \forall i > 0 \qquad (3.71)$$

The relationship ensures that the cost in terms of fixed cost of achieving linear improvements in efficiency increases quadratically. Both marginal and fixed costs span the range of configuration options available so, provided investment in greater efficiency exhibits decreasing returns to scale, the solution will be a series of continuous trade-offs between incrementally different technologies. Where the opposite is true, the plant will either be fully configured or not at all. Formulations with configuration "sweet-spots" will require a more complex approach.

The total cost of operating at a utilisation level, u, is:

$$TC_i(u) = \mathrm{FC}_i^{-} + CFG_i^{2}\left(\mathrm{FC}_i^{+} - \mathrm{FC}_i^{-}\right) + \left(\mathrm{MC}_i^{+} - CFG_i\left(\mathrm{MC}_i^{+} - \mathrm{MC}_i^{-}\right)\right)u \qquad \forall i > 0 \qquad (3.72)$$

To find the optimal configuration for a given utilisation level, u, we minimise $TC_i(u)$ subject to the restriction on the configuration:

$$1 - CFG_i \ge 0 \quad \perp \quad \zeta_i^{cfg} \ge 0 \qquad \forall i > 0 \qquad (3.73)$$

Subject to $CFG_i \ge 0$, we have the following first order conditions defining the optimal configuration for each utilisation level u:

$$2\left(\mathrm{FC}_i^{+} - \mathrm{FC}_i^{-}\right)CFG_i - \left(\mathrm{MC}_i^{+} - \mathrm{MC}_i^{-}\right)u + \zeta_i^{cfg} \ge 0 \quad \perp \quad CFG_i \ge 0 \qquad \forall i > 0 \qquad (3.74)$$

$$1 - CFG_i \ge 0 \quad \perp \quad \zeta_i^{cfg} \ge 0 \qquad \forall i > 0 \qquad (3.75)$$

For $CFG_i \ge 0$ we have:

$$CFG_i = \frac{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)u - \zeta_i^{cfg}}{2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)} \qquad \forall i > 0 \qquad (3.76)$$

Where the optimal configuration is at an interior level, $CFG_i < 1$ and:

$$CFG_i = \frac{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)u}{2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)} \qquad \forall i > 0 \qquad (3.77)$$

Confirming intuition, and in agreement with analysis for other technologies, the higher the target utilisation level, the greater the degree of configuration or investment in operating efficiency is justified, and can be supported. Returning to (3.76), we can re-arrange and solve for u, while substituting $CFG_i = 1$ to identify the utilisation levels at which the limiting technology is the most efficient:

$$u = \frac{2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right) + \zeta_i^{cfg}}{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)} \qquad \forall i > 0 \qquad (3.78)$$

In (3.78) there are a number of utilisation levels that support the limiting configuration, each of which is supported in (3.78) by varying values of the dual variable $\zeta_i^{cfg}$. We are interested in the lowest utilisation level that supports the limiting configuration, at which $\zeta_i^{cfg} = 0$ so that the lowest utilisation level at which the limiting configuration is most efficient is:

$$u_i^* = \frac{2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)}{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)} \qquad \forall i > 0 \qquad (3.79)$$

The optimised cost function is:

$$\begin{aligned} TC_i(u) &= \mathrm{FC}_\mathrm{i}^- + CFG_i^2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right) + \left(\mathrm{MC}_\mathrm{i}^+ - CFG_i\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)\right)u \\ &= \mathrm{FC}_\mathrm{i}^- + \frac{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)^2}{4\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)}u^2 + \left(\mathrm{MC}_\mathrm{i}^+ - \frac{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)^2}{2\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)}u\right)u \qquad \forall i > 0 \qquad (3.80) \\ &= \mathrm{FC}_\mathrm{i}^- + \mathrm{MC}_\mathrm{i}^+ u - \frac{\left(\mathrm{MC}_\mathrm{i}^+ - \mathrm{MC}_\mathrm{i}^-\right)^2}{4\left(\mathrm{FC}_\mathrm{i}^+ - \mathrm{FC}_\mathrm{i}^-\right)}u^2 \end{aligned}$$

As defined in (3.79), the optimised total cost at each utilisation level up to $u_i^*$, can be viewed as a combination of the cost of the least configured technology and a discount relative to that cost, that grows quadratically with the utilisation level being targeted. As described earlier, the interpretation of this total cost function is different to the normal interpretation of a cost function. It does not represent a single technology with quadratic costs, but rather an optimised continuum of infinitesimally different technologies, each with linear costs. Although the interpretation is different, the derivative of this function remains the marginal cost of the marginal plant, and therefore the system price at each utilisation level under perfect competition.

Beyond $u_i^*$, the limiting configuration is most efficient, as no further configuration is available. Where $u_i^* > 1$ is guaranteed, we would not need to consider the configurable version of the technology in our analysis. However, where $u_i^* < 1$ we must consider both the limiting configuration and the quadratic technology, whose possible application must be restricted to utilisation levels $u < u_i^*$. Failure to impose this restriction ignores the limits of configuration and admits the possibility of negative marginal costs as a result of implicitly assuming a never-ending and, in this case, linear ability to improve plant efficiency. Accordingly, the two cost functions that are necessary to describe the optimised total cost of a configurable technology i are:

$$TC_i(u) = \mathrm{FC_i^-} + \mathrm{MC_i^+}u - \frac{\left(\mathrm{MC_i^+} - \mathrm{MC_i^-}\right)^2}{4\left(\mathrm{FC_i^+} - \mathrm{FC_i^-}\right)}u^2 \qquad \forall i > 0, u < u_i^* \qquad (3.81)$$

$$TC_{i+}(u) = \mathrm{FC_i^+} + \mathrm{MC_i^-}u \qquad \forall i > 0, u \geq u_i^* \qquad (3.82)$$

### 3.5.3 Optimal Trade-Offs

Whereas, there is only a single optimal trade-off between two fixed technologies with linear cost structures, interactions between technologies that are in general quadratic, imply the possibility of the total production cost being equal at two utilisation levels in each pairwise comparison. We have partitioned the cost structure of each individual technology, so that those technologies that are configurable are represented as two technologies, unless the configurable portion dominates the limiting version over the utilisation range, $0 \leq u \leq 1$. The utilisation levels corresponding to optimal trade-offs can be inferred by defining the cost difference between two cost functions as a slightly perturbed quadratic and solving for its real roots. We begin by considering the interaction between two quadratic cost functions:

$$\begin{aligned} TC_i(u) - TC_j(u) = {} & \mathrm{FC_i^-} + \mathrm{MC_i^+}u - \frac{\left(\mathrm{MC_i^+} - \mathrm{MC_i^-}\right)^2}{4\left(\mathrm{FC_i^+} - \mathrm{FC_i^-}\right)}u^2 \\ & - \mathrm{FC_j^-} - \mathrm{MC_j^+}u + \frac{\left(\mathrm{MC_j^+} - \mathrm{MC_j^-}\right)^2}{4\left(\mathrm{FC_j^+} - \mathrm{FC_j^-}\right)}u^2 \qquad \forall i, j \neq i \qquad (3.83) \\ & = \mathrm{a_{i,j}}u^2 + \mathrm{b_{i,j}}u + \mathrm{c_{i,j}} \end{aligned}$$

Where:

$$\mathrm{c_{i,j}} = \mathrm{FC_i^-} - \mathrm{FC_j^-} \qquad \forall i, j \neq i \qquad (3.84)$$

$$\mathrm{b_{i,j}} = \mathrm{MC_i^+} - \mathrm{MC_j^+} \qquad \forall i, j \neq i \qquad (3.85)$$

$$\mathrm{a_{i,j}} = \frac{1}{4}\left[\frac{\left(\mathrm{MC_j^+} - \mathrm{MC_j^-}\right)^2}{\mathrm{FC_j^+} - \mathrm{FC_j^-}} - \frac{\left(\mathrm{MC_i^+} - \mathrm{MC_i^-}\right)^2}{\mathrm{FC_i^+} - \mathrm{FC_i^-}}\right] \qquad \forall i, j \neq i \qquad (3.86)$$

The roots of this quadratic define the optimal trade-offs between technologies but there are several pitfalls in the direct implementation of the quadratic formula in this case. We need to consider:

- Linear cost structures, for which $FC_i^+ = FC_i^-$;
- Complex quadratic roots (non-intersection);
- Linear trade-offs, as well as coincidental cases for which $a_{i,j} = 0$; and
- Limiting optimal trade-offs to allowable ranges

The first such issue we encounter is the definition of the quadratic coefficient in each individual cost expression, the difference of which gives us $a_{i,j}$. To avoid numerical difficulties when technology i has a linear cost structure we define $a_i$ as follows:

$$a_i = \frac{\left(MC_i^+ - MC_i^-\right)^2}{4\left(FC_i^+ - FC_i^- + \left(1 - a_i^{cfg}\right)\varepsilon\right)} \qquad \forall i \qquad (3.87)$$

Where the cost structure is linear, although the denominator is perturbed the expression evaluates as zero, which is the desired outcome. Similarly, when the cost structure is quadratic, the denominator is not perturbed, and the expression evaluates correctly on this occasion also. We define $a_{i,j}$ using the definition (3.87):

$$a_{i,j} = a_j - a_i \qquad \forall i, j \neq i \qquad (3.88)$$

To reflect the multiplicity of roots, we define $u^e_{i,j,1}$ and $u^e_{i,j,2}$ to be the roots of (3.83), corresponding to the '+' and '-' variations of the standard quadratic formula respectively:

$$u^e_{i,j,(1,2)} = \frac{-b_{i,j} \pm \sqrt{b_{i,j}^2 - 4a_{i,j}c_{i,j}}}{2a_{i,j}} \qquad (3.89)$$

As shown in (3.89), the roots are exogenous and can be precompiled from exogenous cost data. These could be used directly as candidates for system-wide critical utilisation levels however due to capacity inflexibility and energy limits, cost structures or more likely the values of imputed costs, are often redefined endogenously. Therefore, without specifying any particular source or cause for endogeneity, or addressing the particular formulation adjustments required to specify capacity or energy limits in this context, we proceed to develop the formulation on the basis that $a_{i,j}$, $b_{i,j}$, and $c_{i,j}$ are endogenous.

The second issue arising from this approach is the need to avoid complex roots. The roots of (3.83) are real whenever $b_{i,j}^2 - 4a_{i,j}c_{i,j} \geq 0$, however whenever $b_{i,j}^2 - 4a_{i,j}c_{i,j} < 0$ they are complex and reflect complete dominance of one technology over another, thereby depriving our approach of an optimal trade-off level, at any utilisation level, whether feasible or not. Our approach partitions the range of $b_{i,j}^2 - 4a_{i,j}c_{i,j}$ into two non-negative ranges, and applies asymmetric penalties to force any complex roots to values outside the utilisation range under consideration, $0 \leq u \leq 1$. The following complementarity conditions partition the range:

$$\zeta_{i,j}^{1} - {b_{i,j}}^{2} + 4a_{i,j}c_{i,j} \geq 0 \quad \perp \quad \zeta_{i,j}^{1} \geq 0 \qquad \forall i, j \neq i \tag{3.90}$$

$${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} - \zeta_{i,j}^{1} + \zeta_{i,j}^{2} \geq 0 \quad \perp \quad \zeta_{i,j}^{2} \geq 0 \qquad \forall i, j \neq i \tag{3.91}$$

Where ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} > 0$ then, from (3.90) we have $\zeta_{i,j}^{1} = {b_{i,j}}^{2} - 4a_{i,j}c_{i,j}$. From (3.91), as a result of terms cancelling, we have $\zeta_{i,j}^{2} = 0$. Where ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} < 0$ then, from (3.90), $\zeta_{i,j}^{1} = 0$, and $\zeta_{i,j}^{2} = 4a_{i,j}c_{i,j} - {b_{i,j}}^{2}$, from (3.91). Finally, if we consider the case where ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} = 0$, we have $\zeta_{i,j}^{1} = \zeta_{i,j}^{2} = 0$. By implication we have $\zeta_{i,j}^{1} \geq 0 \perp \zeta_{i,j}^{2} \geq 0$, so that in each case $\zeta_{i,j}^{1} + \zeta_{i,j}^{2} = \left| {b_{i,j}}^{2} - 4a_{i,j}c_{i,j} \right|$, a sum that would suffice as a surrogate for ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j}$ except that where ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} < 0$, simply reversing the sign would create optimal trade-offs that do not exist. To be clear, this would not be injurious to the prospect of a solution being found as it would merely result in the calculation of additional market clearance corresponding to the additional utilisation level. Depending on the complexity of the market clearance, it may well be computationally advantageous to perform that calculation, relative to applying the scaling factor approach we discuss next.

As an alternative, we can apply a scaling factor when undesirable situations such as ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} < 0$ arise. This is a general approach that can be used in a variety of situations to create measures that avoid numerical problems, or to reflect logical stipulations. We replace ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j}$ with the following:

$$\zeta_{i,j}^{1} + \frac{\zeta_{i,j}^{2}}{\varepsilon} \qquad \forall i, j \neq i \tag{3.92}$$

When ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} \geq 0$ we have $\zeta_{i,j}^{2} = 0$, yielding the desired unaltered outcome:

$$\zeta_{i,j}^{1} = {b_{i,j}}^{2} - 4a_{i,j}c_{i,j} \qquad \forall i, j \neq i \tag{3.93}$$

When ${b_{i,j}}^{2} - 4a_{i,j}c_{i,j} < 0$ we have $\zeta_{i,j}^{1} = 0$, and (3.92) yields:

$$\frac{\zeta_{i,j}^{2}}{\varepsilon} = -\frac{\left( {b_{i,j}}^{2} - 4a_{i,j}c_{i,j} \right)}{\varepsilon} \qquad \forall i, j \neq i \tag{3.94}$$

Provided $\varepsilon$ is chosen to be sufficiently small, any non-trivial complex roots will be transformed into real, and large, utilisation levels that will be ignored in the same fashion as other optimal trade-offs outside the appropriate range $0 \leq u_{i,j}^{e,(1,2)} \leq 1$. Where $\varepsilon$ is not chosen appropriately, successful numerical evaluation will still occur, however the expression will generate spurious optimal trade-offs in these cases. The error resulting from the introduction of scaling factors into the selection of technologies can be made arbitrarily small by selecting progressively smaller values for $\varepsilon$, so that the only technologies capable of being confused by the complementarity conditions are practically

identical in terms of their cost structure, and could legitimately be considered a single technology. We can substitute (3.92) into (3.89):

$$u^{e}_{i,j,(1,2)} = \frac{-b_{i,j} \pm \sqrt{\zeta^{1}_{i,j} + \frac{\zeta^{2}_{i,j}}{\varepsilon}}}{2a_{i,j}} \qquad \forall i, j \neq i \qquad (3.95)$$

From (3.89), whenever the denominator is zero the optimal trade-offs are undefined. Therefore, we require a strategy to further modify (3.89) to avoid computational issues when $a_{i,j} = 0$. This situation arises when either when the comparison is between linear cost structures, or by coincidence where two technologies share the same quadratic cost coefficient. It is tempting to simply perturb the denominator in a simple fashion so that $2a_{i,j}$ becomes $2a_{i,j} \pm \varepsilon$. But this would not resolve the general numerical issue, as zero valued denominators would still be possible. To ensure the denominator is non-zero, we adopt a similar approach as before and partition the variable $a_{i,j}$ into two non-negative complementary components, $\zeta^{3}_{i,j}$ and $\zeta^{4}_{i,j}$, although this time our goal is to exclude zero, rather than rule out the negative half-space as before:

$$\zeta^{3}_{i,j} - a_{i,j} \geq 0 \quad \perp \quad \zeta^{3}_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (3.96)$$

$$a_{i,j} - \zeta^{3}_{i,j} + \zeta^{4}_{i,j} \geq 0 \quad \perp \quad \zeta^{4}_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (3.97)$$

Where $a_{i,j} > 0$, then from (3.96) we have $\zeta^{3}_{i,j} = a_{i,j}$. From (3.97), as a result of terms cancelling, we have $\zeta^{4}_{i,j} = 0$. Where $a_{i,j} < 0$ then, from (3.96), $\zeta^{3}_{i,j} = 0$ and $\zeta^{4}_{i,j} = -a_{i,j}$ from (3.97). Finally, if we consider the case where $a_{i,j} = 0$, we have $\zeta^{1}_{i,j} = \zeta^{2}_{i,j} = 0$.

Although we could utilise separate perturbation variables and there could be specific reasons related to problem scaling and numerical efficiency guiding that choice, the setting of $\varepsilon$ to a sufficiently low level is the relevant requirement. We abuse notation and re-use $\varepsilon$ as a perturbation term noting that, in whichever context it is used, a lower value represents an improvement in solution accuracy. The following expression produces a perturbed denominator that can be made arbitrarily close to the true value:

$$\zeta^{3}_{i,j} - \zeta^{4}_{i,j} + \left(1 - \frac{\zeta^{3}_{i,j} - \zeta^{4}_{i,j}}{\zeta^{3}_{i,j} - \zeta^{4}_{i,j} + \varepsilon}\right)\varepsilon \qquad \forall i, j \neq i \qquad (3.98)$$

Where $a_{i,j} = 0$:

$$\left(1 - \frac{0}{\varepsilon}\right)\varepsilon \rightarrow \varepsilon \; as \; \varepsilon \rightarrow 0 \qquad (3.99)$$

Where $a_{i,j} > 0$, or $a_{i,j} < 0$ we respectively have:

$$\zeta_{i,j}^{3}+\left(1-\frac{\zeta_{i,j}^{3}}{\zeta_{i,j}^{3}+\varepsilon}\right)\varepsilon\rightarrow\zeta_{i,j}^{3}=a_{i,j}\ as\ \varepsilon\rightarrow 0 \qquad \forall i,j\neq i \qquad (3.100)$$

$$-\zeta_{i,j}^{4}+\left(1-\frac{-\zeta_{i,j}^{4}}{-\zeta_{i,j}^{4}+\varepsilon}\right)\varepsilon\rightarrow-\zeta_{i,j}^{4}=a_{i,j}\ as\ \varepsilon\rightarrow 0 \qquad \forall i,j\neq i \qquad (3.101)$$

Whereas a simple perturbation of the denominator merely shifts the problem of a zero denominator along the axis, this approach drives the estimates of $a_{i,j}$ away from zero in the appropriate direction. As a consequence of the expression defined in (3.98) being close to zero, the root will be defined but the utilisation level corresponding to it will be outside the desired range. Substituting (3.98) into (3.95) we have (3.102), an expression that, given an appropriate choice of perturbation, defines the utilisation levels corresponding to technological trade-offs as defined by the roots of the quadratic (3.83):

$$u_{i,j,(1,2)}^{e}=\frac{-b_{i,j}\pm\sqrt{\zeta_{i,j}^{1}+\frac{\zeta_{i,j}^{1}}{\varepsilon}}}{2\left(\zeta_{i,j}^{3}+\zeta_{i,j}^{4}+\left(1-\frac{\zeta_{i,j}^{3}+\zeta_{i,j}^{4}}{\zeta_{i,j}^{3}+\zeta_{i,j}^{4}+\varepsilon}\right)\varepsilon\right)} \qquad \forall i,j\neq i \qquad (3.102)$$

Whenever those quadratic roots they are real, they are evaluated. Where they are not, the evaluation of (3.102) yields utilisation levels outside the range of utilisation levels we are interested in. This resolves the numerical issue where $a_{i,j}=0$ as a result of quadratic cost structures coinciding. Unfortunately, when we consider the optimal trade-off between linear cost structures, this approach is not appropriate as it will not determine the correct optimal trade-off between linear technologies, which is given by $u_{i,j}^{e}=-c/b$. To see this, consider (3.89) as $a_{i,j}\rightarrow 0$:

$$\begin{aligned} u_{i,j,(1,2)}^{e} &= \frac{-b_{i,j}\pm\sqrt{{b_{i,j}}^{2}-4a_{i,j}c_{i,j}}}{2a_{i,j}+\varepsilon} \\ &= \frac{-b_{i,j}+\sqrt{{b_{i,j}}^{2}}}{\varepsilon}\ or\ \frac{-b_{i,j}-\sqrt{{b_{i,j}}^{2}}}{\varepsilon} \\ &= 0\ or\ \frac{-2b_{i,j}}{\varepsilon} \end{aligned} \qquad \forall i,j\neq i \qquad (3.103)$$

The non-zero solution of (3.103) is a value unrelated to the desired optimal trade-off between linear cost structures. The reverse is also true, as $u_{i,j}^{e}=-c/b$ will not generate the correct solution in the case of coincident quadratic cost-coefficients. Given the expression of roots in (3.102) and $u_{i,j}^{e}=-c/b$, a selection must be made, depending on whether the original cost structures being compared were both linear or not. We must choose between these expressions and, where the quadratic interaction is appropriate, identify both potential roots. The following expressions define the optimal trade-off between technologies i and j, $\forall i,j\neq i$ that we desire by developing an appropriately weighted convex combination of each possibility, using weights assigned by functions of the variable

$\mathrm{a}^{\mathrm{cfg}}$ that is set to unity for configurable technologies and zero to limiting or non-configurable technologies:

$$u^{e}_{i,j,1}=\left(1-\frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}+\varepsilon}\right)\left(\frac{-c}{b}\right)+\left(\frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}+\varepsilon}\right)\left[\frac{-b_{i,j}+\sqrt{\zeta^{1}_{i,j}+\dfrac{\zeta^{1}_{i,j}}{\varepsilon}}}{2\left(\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\left(1-\dfrac{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}}{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\varepsilon}\right)\varepsilon\right)}\right] \tag{3.104}$$

$$u^{e}_{i,j,2}=\left(1-\frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}+\varepsilon}\right)\left(\frac{-c}{b}\right)+\left(\frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}+\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}+\varepsilon}\right)\left[\frac{-b_{i,j}-\sqrt{\zeta^{1}_{i,j}+\dfrac{\zeta^{1}_{i,j}}{\varepsilon}}}{2\left(\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\left(1-\dfrac{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}}{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\varepsilon}\right)\varepsilon\right)}\right] \tag{3.105}$$

Where at least one technology has a quadratic cost structure, then either $\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}=1$ or $\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}=1$, or both. As $\varepsilon\to 0$ we have:

$$u^{e}_{i,j,(1,2)}\to\frac{-b_{i,j}\pm\sqrt{\zeta^{1}_{i,j}+\dfrac{\zeta^{1}_{i,j}}{\varepsilon^{1}}}}{2\left(\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\left(1-\dfrac{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}}{\zeta^{3}_{i,j}+\zeta^{4}_{i,j}+\varepsilon^{2}}\right)\varepsilon^{2}\right)}\qquad\forall i,j\neq i \tag{3.106}$$

As already described, this expression has been adjusted to accommodate the situation where $a_{i,j}=0$ and $b_{i,j}{}^{2}-4a_{i,j}c_{i,j}<0$. Where each technology has a linear cost structure, then $\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}=\mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}=0$ and no such correction is needed:

$$u^{e}_{i,j,1}=u^{e}_{i,j,2}=\frac{-c}{b} \tag{3.107}$$

There is a duplication of optimal trade-offs in this case, but these definitions only seed values to the complementarity constraints that define critical system-wide utilisation levels so that, beyond the unnecessary computation involved, the duplication of optimal trade-offs is not a problem in this structure.

In Chapter 2, we defined optimal trade-offs as follows:

$$\mathrm{FC}_{i}-\mathrm{FC}_{j}-(\mathrm{MC}_{j}-\mathrm{MC}_{i})u^{e}_{i,j}+\eta_{i,j}\geq 0\quad\perp\quad u^{e}_{i,j}\geq 0\qquad\forall i,j\neq i \tag{3.108}$$

$$1-u^{e}_{i,j}\geq 0\quad\perp\quad\eta_{i,j}\geq 0\qquad\forall i,j\neq i \tag{3.109}$$

(3.108) and (3.109) implicitly defined optimal trade-offs. To accommodate investment in configurable technologies, we replace (3.108) with an explicit definition of the optimal technological trade-offs, $\forall i,j\neq i$:

$$u^{e}_{i,j,v} - \left(1 - \frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}} + \mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}} + \mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}} + \varepsilon}\right)\frac{-c}{b} - \left(\frac{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}} + \mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}}}{\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}} + \mathrm{a}^{\mathrm{cfg}}_{\mathrm{j}} + \varepsilon}\right)\left[\frac{-b_{i,j} + \sqrt{\zeta^{1}_{i,j} + \frac{\zeta^{1}_{i,j}}{\varepsilon^{1}}}}{2\left(\zeta^{3}_{i,j} + \zeta^{4}_{i,j} + \left(1 - \frac{\zeta^{3}_{i,j} + \zeta^{4}_{i,j}}{\zeta^{3}_{i,j} + \zeta^{4}_{i,j} + \varepsilon^{2}}\right)\varepsilon^{2}\right)}\right] + \eta_{i,j,v} \geq 0$$

$$\perp \quad u^{e}_{i,j,v} \geq 0 \qquad \forall i, j \neq i, v \qquad (3.110)$$

As discussed earlier, in a pairwise comparison between the configurable and limiting versions of a particular technology, the quadratic technology, which by definition represents the efficient frontier of configuration options, will match all individual expressions of that particular technology within the allowable range, and dominate them outside of that range. But it is clear that outside the range defined for it, the imputed quadratic cost structure is not defined as, if unchecked, the quadratic portion of the cost structure may generate further invalid optimal trade-offs with other technologies, and eventually dominate all technologies. To restrict the bounds of validity, the roots must also be limited to the appropriate utilisation range, which in general is $0 \leq u^{e}_{i,j,(1,2)} \leq 1$. We must restrict each notionally quadratic technology to a maximum utilisation range corresponding to the most capital-intensive configuration available. That maximum utilisation level is given by:

$$u^{*}_{i} = \frac{2\left(\mathrm{FC}^{+}_{\mathrm{i}} - \mathrm{FC}^{-}_{\mathrm{i}}\right)}{\left(\mathrm{MC}^{+}_{\mathrm{i}} - \mathrm{MC}^{-}_{\mathrm{i}}\right)} \qquad \forall i \qquad (3.111)$$

We can state a general bound whose value depends on the technology under consideration:

$$u^{e,+}_{i,j,v} = \mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}\left[\frac{2\left(\mathrm{FC}^{+}_{\mathrm{i}} - \mathrm{FC}^{-}_{\mathrm{i}}\right)}{\left(\mathrm{MC}^{+}_{\mathrm{i}} - \mathrm{MC}^{-}_{\mathrm{i}}\right)}\right] + \left(1 - \mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}\right) \qquad \forall i \qquad (3.112)$$

Accordingly, we replace (3.109) with the following complementarity condition which sets the upper limit on utilisation for technology i to either unity, in the case where technology is linear technology, or the limit described in (3.111), in the case where technology i represents the range in which an optimised installation of a configurable technology is described by a non-linear cost structure:

$$\mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}\left[\frac{2\left(\mathrm{FC}^{+}_{\mathrm{i}} - \mathrm{FC}^{-}_{\mathrm{i}}\right)}{\left(\mathrm{MC}^{+}_{\mathrm{i}} - \mathrm{MC}^{-}_{\mathrm{i}}\right)}\right] + \left(1 - \mathrm{a}^{\mathrm{cfg}}_{\mathrm{i}}\right) - u^{e}_{i,j} \geq 0 \quad \perp \quad \eta_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (3.113)$$

### 3.5.4 Critical Utilisation Levels

The approach described in Chapter 2 sequentially examines a set of optimal trade-offs to determine the critical trade-offs. In that case, the interaction between two linear cost functions yields a maximum of one intersection. While the interaction between two linear technologies generates a single trade-off, when a pairwise comparison involves consideration of a set of technological configurations with quadratic trade-offs, we must contemplate the possibility of two optimal trade-offs. Dual optimal trade-offs reflect the reality that, at least in a pairwise sense and ignoring utilisation bounds, the more configurable technology will have discrete operational niches or efficient operating zones, while the

less configurable technology will be most efficient only in intermediate roles. Naturally that relationship is potentially spoiled by utilisation bounds, and cannot be extended beyond pairwise comparison as one or other set of configurations may be dominated by other technologies.

Several adjustments to the process of selecting critical utilisation levels must be made to accommodate the dual intersections and non-constant marginal costs that come with quadratic cost structures. Most fundamentally, we must adjust the selection variable $z_{j,n}$ itself. We require the selection variable to include a further dimension, indexed by v={1,2} to reference the two possible intersections between the current technology i, and a candidate for selection, technology j. With $z_{j,n}$ becoming $z_{j,n,v}$, the selection constraint becomes:

$$u_n^e - \sum_j z_{j,n,v} \sum_{i,v} \mathrm{z}_{\mathrm{i,n\text{-}1,v}} \mathrm{u}^{\mathrm{e}}_{\mathrm{i,j,v}} = 0 \qquad :\psi_n^0 \qquad \forall n > 0 \qquad (3.114)$$

The following set of amended original constraints collectively ensure that a single choice is made at each stage n, and require no further adjustment other than to extend the dimensionality to incorporate multiple optimal trade-offs per technology:

$$\sum_{j,v} z_{j,n,v} - 1 \geq 0 \qquad :\psi_n^1 \qquad \forall n \qquad (3.115)$$

$$\sum_{j,v} z_{j,n,v}{}^2 - 1 \geq 0 \qquad :\psi_n^5 \qquad \forall n \qquad (3.116)$$

$$z_{j,n,v} \geq 0 \qquad \forall j,n,v \qquad (3.117)$$

We have noted that when we allow quadratic technologies, each may intersect twice with another technology. However, it does not follow that these technologies may have a limit of two intersections with the screening curve lower envelope. In general, they may have many, which suggests that highly configurable technologies might be capable of filling several niche generation roles. Nevertheless, the determination of critical utilisation levels is based on selecting from intersections with a particular technology, enabling us to limit our exploration of critical utilisation levels to just two intersections per pairwise technological combination.

### 3.5.5 Deriving the PDC

The optimal PDC can be constructed by interpolating the marginal cost between the beginning and end of each critical utilisation. We have indexed each segment of the PDC by n=1,..N, where n=1 corresponds to shortage events and prices, and n=N corresponds to base load operations. Using the derivative of the optimised total cost function (3.80), the marginal cost as defined at the start of PDC segment n is given by:

$$MC_n^{start} = \sum_{j,v} z_{j,n-1,v} \left( \mathrm{MC}_{\mathrm{j}}^{+} - \frac{\mathrm{a}_{\mathrm{j}}^{\mathrm{cfg}}}{2} \left[ \frac{\left( \mathrm{MC}_{\mathrm{i}}^{+} - \mathrm{MC}_{\mathrm{i}}^{-} \right)^2}{\mathrm{FC}_{\mathrm{i}}^{+} - \mathrm{FC}_{\mathrm{i}}^{-}} \right] u_n^e \right) \qquad \forall n > 0 \qquad (3.118)$$

(3.118) defines the marginal cost of the marginal technology at $u_n^e$. At stage n-1 this (perhaps configurable) technology enters the algorithm to define the lower envelope of the screening curve at step n-1 and is marginal across the utilisation range defined by $\left\{u_{n-1}^e, u_n^e\right\}$. The marginal cost corresponding to the initial operation is therefore evaluated at $u_n^e$, which corresponds to the load level at which technology j would commence generation. The presence of the configuration indicator $a_j^{cfg}$ ensures that numerical difficulties are avoided when considering limiting or non-configurable technologies.

We can analogously define marginal costs at the other end of the marginal operating range in focus. This is given by:

$$MC_n^{end} = \sum_{j,v} z_{j,n-1,v} \left( \mathrm{MC}_j^+ - \frac{a_i^{cfg}}{2} \left[ \frac{\left(\mathrm{MC}_i^+ - \mathrm{MC}_i^-\right)^2}{\mathrm{FC}_i^+ - \mathrm{FC}_i^-} \right] u_{n-1}^e \right) \qquad \forall n > 0 \qquad (3.119)$$

The same determination is made in (3.119), although this time it is evaluated at $u_{n-1}^e$, which corresponds to the point where technology j reaches full capacity and becomes inframarginal.

When the technology has a linear cost structure then we have $MC_n^{start} = MC_n^{end}$. For pseudo-technologies with linearly varying marginal costs, the assumption of linear adjustment in utilisation guarantees the PDC can be generated with a simple linear interpolation between $\left(MC_n^{start}, u_n^e\right)$ and $\left(MC_n^{end}, u_{n-1}^e\right)$.

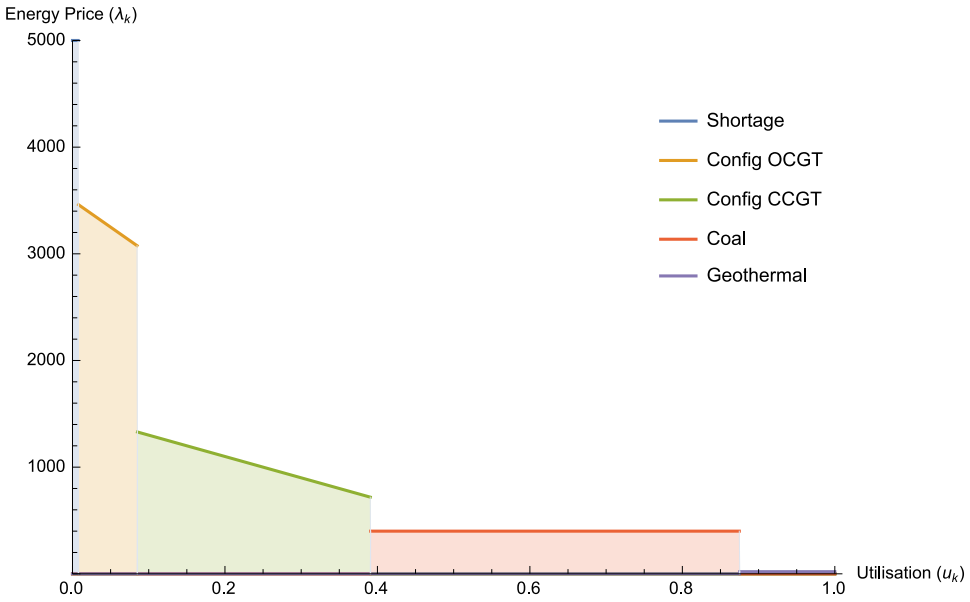



**Figure 26: PDC with Configurable Technologies**

As shown in Figure 26, the PDC may have several possible features. Where technologies are linear we have the standard case of constant segments in the PDC, corresponding to a series of technologies offering at a constant marginal cost. Where a non-linear technology is configurable its cost structure is

quadratic, which corresponds to linear marginal costs that, under perfect competition, translate to linear prices across that utilisation range. In Figure 26, the configurable technologies shown are OCGT and CCGT. Note that with discrete capital cost ranges there still remains discrete changes in pricing between technologies. Where capital cost ranges overlap one optimised technology will merge into another, and if they are both marginal, the adjustment in pricing will be continuous, albeit at different rates when viewed from the perspective of utilisation. The PDC is therefore potentially a succession of non-linear technologies being price setting, each of which are eventually superseded by either a linear technology or another non-linear technology and vice versa, resulting in linear and piecewise constant segments.

The derivation of the PDC suggests a number of alterations to the market and investment equilibrium conditions. By defining utilisation levels to correspond to optimal trade-offs we generate operating ranges in which configurable technologies are marginal in the same fashion as with conventional technologies. However, by virtue of the fact capacity of a configurable technology is not homogeneous, these technologies are not readily interpreted as conventional technologies. To clarify the performance of individual tranches of configuration, we must consider the capacity of each configuration range separately. Sometimes these tranches may be contiguous but other times they will not, as it is possible for a configurable technology i to be employed in a number of distinct roles. Accordingly, for each operating range, r, there is a distinct technology i(r), with a total capacity $CAP_{i(r)}$. At this point we need to clarify that we still maintain the broader technological characteristics for each configurable technology so that the optimal trade-offs involving these technologies are able to be determined, but we are using "sub-technologies", indexed by i(r), to correspond to the tranches that are actually installed. Accordingly, in what follows, several constraints will not apply to the broader configurable technology, and instead relate to the sub-technologies defined by actual installation.

The efficiency of the market clearing process is governed by the basic dual relationship between market prices and the marginal cost of generation. The standard formulation of this condition is unaltered, except for the domain restriction that means it does not apply to configurable technologies.

$$-\lambda_r + \mathrm{MC}_{\mathrm{i}}^{+} + \varphi_{i,r}^{+} \geq 0 \quad \perp \quad GEN_i \geq 0 \qquad \forall i \notin C, r \tag{3.120}$$

These technologies are accounted for explicitly. As shown below the form of marginal cost is significantly more complex, and is comprised of the derivative of the optimised cost function:

$$-\lambda_r + \left( \mathrm{MC}_{\mathrm{i(r)}}^{+} - \frac{1}{2}\left[ \frac{\left(\mathrm{MC}_{\mathrm{i(r)}}^{+} - \mathrm{MC}_{\mathrm{i(r)}}^{-}\right)^2}{\mathrm{FC}_{\mathrm{i(r)}}^{+} - \mathrm{FC}_{\mathrm{i(r)}}^{-}} \right] u_r \right) + \varphi_{i(r),r}^{+} \geq 0 \quad \perp \quad GEN_{i(r)} \geq 0 \qquad \forall i(r), r \tag{3.121}$$

Because the cost structure of each configurable technology is optimised, then by definition the merit order as it relates to specific configurations of a particular technology is respected across any utilisation range in which such a technology is marginal.

The standard generation constraint for conventional technologies remains the same and we add an equivalent constraint that applies to all configurable technologies i(r):

$$CAP_i - GEN_{i,r} \geq 0 \quad \perp \quad \varphi^+_{i,r} \geq 0 \qquad \forall i > 0, i \notin C, r \quad (3.122)$$

$$CAP_{i(r)} - GEN_{i(r),r} \geq 0 \quad \perp \quad \varphi^+_{i(r),r} \geq 0 \qquad \forall i(r), r \quad (3.123)$$

The market clearing condition is given by:

$$\sum_{i \notin C} GEN_{i,r} + \sum_{i(r)} GEN_{i(r)} - L_r = 0 \quad \perp \quad \lambda_{r,t} \; free \qquad \forall r, t \quad (3.124)$$

Finally, we turn to the investment condition. We note that the profitability of each configuration of the technology i(r) is identical by construction, as the optimisation of the total cost function explicitly balances the rate of change in fixed costs and variable costs. Returning to the PDC definition, it is clear that the profitability of each technology involves a combination of linear and constant pricing segments, and therefore the profitability defined in (2.21) does not describe the profitability of each technology across all utilisation ranges. Whether a technology is configurable or not, we must define the profitability of each technology in the operating range defined by $\{u_r, u_{r+1}\}$.

The profitability at $u_{r+1}$ is given by the variable $\varphi^+_{i,r+1}$, or $\varphi^+_{i(r),r+1}$. As we have optimised the utilisation levels corresponding to operating ranges, each operating range has a single marginal technology. Where that technology exhibits constant marginal costs, the profitability of the technology is given by multiplying $\varphi^+_{i,r+1}$ by the width of the utilisation range, $u_{r+1} - u_r$. Where the marginal technology is actually a configurable technology and represents a continuously adjusting marginal cost this definition is not appropriate. It is also clear from Figure 26 that we cannot simply use the next level of profitability, $\varphi^+_{i,r}$, to determined average profitability over the operating range, as this in general will not reflect the price adjustment over a configuration range. We might actually define the PDC directly using the variables $MC_n^{start}$ and $MC_n^{end}$ from (3.118) and (3.119) but instead we elect to define it using perturbed market clearances.

Re-using the perturbation term $\varepsilon$ we define utilisation levels, $u_r^\varepsilon = u_r + \varepsilon$, and construct alternative market clearances using identical constraint structures as shown from (3.120) to (3.124). The load corresponding to the utilisation level $u_r^\varepsilon$ is given by:

$$L_r^\varepsilon = L_r - \frac{L_r - L_{r+1}}{u_{r+1} - u_r}\varepsilon \qquad \forall r \quad (3.125)$$

By solving the system at these perturbed levels $\{u_r^\varepsilon, L_r^\varepsilon\}$ we capture market prices and profitability, respectively $\lambda_r^\varepsilon$ and $\varphi^{\varepsilon,+}_{i,r}$, $\varphi^{\varepsilon,+}_{i(r)}$. We can then define the average profitability within a utilisation range using the perturbed market clearances. Without loss of generality, and adjusting for the perturbed operating range length, the average profitability of a conventional technology within an operating range is defined as:

$$\text{AvgProfit} = \varphi^{+}_{i,r+1} + \frac{1}{2}\left(\varphi^{+,\varepsilon}_{i,r} + \varphi^{+}_{i,r+1}\right)\left(\frac{u_{r+1} - u_r}{u_{r+1} - u^{\varepsilon}_r}\right) \qquad \forall i \qquad (3.126)$$

Accordingly, the investment constraint for a conventional technology is:

$$\text{FC}_\text{i} - \sum_{r<R}\left[\varphi^{+}_{i,r+1} + \frac{1}{2}\left(\varphi^{+,\varepsilon}_{i,r} + \varphi^{+}_{i,r+1}\right)\left(\frac{u_{r+1} - u_r}{u_{r+1} - u^{\varepsilon}_r}\right)\right]\left(\text{u}_{r+1} - \text{u}_r\right) \geq 0 \perp CAP_i \geq 0 \qquad \forall i \notin C \qquad (3.127)$$

For a configurable technology, it is not the case that fixed costs are constant. Optimised fixed costs depend on the utilisation level of the technology concerned. We have:

$$\text{FC}_\text{i(r)} - \sum_{r<R}\left[\varphi^{+}_{i(r),r+1} + \frac{1}{2}\left(\varphi^{+,\varepsilon}_{i(r),r} + \varphi^{+}_{i(r),r+1}\right)\left(\frac{u_{r+1} - u_r}{u_{r+1} - u^{\varepsilon}_r}\right)\right]\left(\text{u}_{r+1} - \text{u}_r\right) \geq 0 \perp CAP_{i(r)} \geq 0 \;\; \forall i(r) \qquad (3.128)$$

Where:

$$\text{FC}_\text{i(r)} = \text{FC}^{\text{-}}_\text{i(r)} + \frac{1}{4}\left(\frac{\left(\text{MC}^{+}_\text{i} - \text{MC}^{\text{-}}_\text{i}\right)^2}{\text{FC}^{+}_\text{i} - \text{FC}^{\text{-}}_\text{i}}\right)u_r^2 \qquad \forall i \qquad (3.129)$$

Although the investment constraint only contemplates a single incarnation of each configurable technology within a tranche, it is the case that if this technology returns a profit of precisely zero, then as these are optimised so will all of the other configurations available in that operating range. As we have stated, traditional technologies make no profit while they are marginal generator. When viewed as a single technology, it appears that configurable technologies are profitable while marginal, however the apparent contradiction is based entirely on an incorrect interpretation of the cost curve and the technology. For each atomistic incarnation of a particular technology, the technology will make zero profit while it is the marginal generator.

## *3.6 Summary*

In this chapter, we have endeavoured to provide a sample of some extensions to the framework. Given the number of possible avenues that could be explored a comprehensive treatment of all is not possible.

Generalising technological cost structures enables consideration of more realistic plant economics, and makes clear the nature of investment in a technology is akin to purchasing a portfolio of plants, where the portfolio weightings are defined by the engineered capability of the technology. Other possibilities for linked investments are no doubt possible, and would exhibit similar forms of trade-off using imputed values of the individual components of the decision.

Sections 3.3.2 and 3.3.3 were devoted to capacity flexibility; first in general and then specifically addressing mothballing and retirement of existing capacity. We showed the nature of capacity of restrictions was to change the optimal trade-off away from that which would naturally occur, with the difference in capacity value summarising the cost of the restriction, or the value of the permit or right that would release the investor from the restriction. We clarified the status of mothballing and reinstatement as operational choices, which offer the capacity investor the possibility,

of improving their returns. In individual periods, the reinstatement of mothballed capacity can significantly affect returns and utilisation of other technologies. This is a particularly relevant issue in markets where older technologies traditionally used in peaking roles are being disrupted by mainly renewable technologies, with less reliability. By diminishing the opportunity to earn profits at these key times, the economic justification for retaining that capacity, and the system reliability may also be diminished. In terms of the framework we present, the additional operating strategy modifies the investment constraint. While sub-period profitability is affected, the definitions are not, and no change is needed to the framework in respect of these.

Energy limits were explored in both the deterministic and stochastic cases. The basic principle of valuing energy according to its opportunity cost is well known. In both cases, optimal trade-offs are modified by adjustments in imputed fuel costs. This requires the substitution of fuel costs with a dual variable representing the opportunity cost of fuel, to ensure that optimal trade-offs are struck using an economic valuation of marginal cost. That opportunity cost is restricted in several ways in our framework. Firstly, we considered endogenous energy limits, as we envisage an operator acquiring take or pay contracts. The value of these contracts is assumed exogenous and that implies a lower bound on the opportunity cost of fuel. As we graduate from single fuel arrivals, to sub-period fuel arrivals we reiterate the work of others in describing the influence of storage constraints. These prevent the unfettered allocation of fuel across sub-periods and result in diverging opportunity costs between periods separated by a binding storage constraint.

Finally, we presented a novel, if not easily formulated, approach to extending the screening curve analysis to address configurable technologies. Our approach was not simply a matter of optimising the configuration of a single technology choice. Instead, we contemplated the optimal configuration range for a particular technology. The effect is to introduce non-linearity into the screening curve diagram but also to fundamentally change what is meant by a technology and capacity in that context.

We define technologies by their limiting fixed and variable costs pairs, and the rate of relative adjustment between them. There must be decreasing returns to scale for investment in efficiency if there is to be a range of interior trade-offs within the variants of the technology. Otherwise the solution would only involve one or other limiting variant of the technology. The assumption of decreasing returns to scale for higher specification configurations allows us to define a piecewise non-linear cost structure that represents the optimised cost structure for that technology.

Having optimised the intra-technology cost structure, we introduce this to the screening curve analysis as a notional quadratic technology, which by virtue of its non-linearity and the potential non-linearity of other technologies leads us to consider multiple optimal trade-offs. The economic implications of this are clear. As opposed to linear cost structures, non-linear cost structures permit the existence of more than one distinct operating range or niche for a single technology.

In this particular example, we hypothesise quadratic adjustment of fixed costs in response to linear improvements in efficiency, so that the PDC contains piecewise constant and linear segments. The definition of profitability for configurable technologies differs from that of conventional technologies. While more complicated, once it is understood that as all configurations in the same

operating range share represent equal total cost on account of the prior optimisation of cost structures, we can define the investment constraint in terms of a single version of the technology, which in this case was the least efficient in terms of variable cost.

This proved to be a highly technical and original exercise that we do not expect to be implemented but nevertheless yields some insight into the extensibility of complementarity formulations. The complementarity formulation involved significant changes to the definition of optimal trade-offs, as there are now two pairwise trade-offs possible between each technology. This involved the development of complementarity conditions to remove imaginary roots, that correspond to non-intersection of technologies from consideration. Independently of our framework, the generalisation of cost structures, and the introduction of capacity and energy limits is standard fare and can be included in a conventional optimisation formulation. However, our contribution was to integrate them with our framework, and to our knowledge that approach is new. In the case of the configurable technologies, we are unaware of a similar approach in the context of screening curve analysis, even without considering the wider framework we develop. Certainly, the full integration of configurable technologies in our framework is novel. We are also unaware of the application of complementarity conditions to filter complex solutions to quadratic equations.

# 4 ENDOGENOUS LOAD & RELIABILITY

## *4.1 Introduction*

In this chapter we continue to illustrate the extensibility of the framework. We do so in this case by revisiting variability that can be represented proportionally within the LDC. We introduce endogeneity to the LDC directly, in the form of demand response, and indirectly, in the form of reliability and intermittent generation.

In Section 4.2, we begin by considering modelling options for demand response in our framework with a view to developing a more consistent view of demand response that is typical in the literature. Demand response is divided into short term demand response of the kind that is physically feasible on the timescale of an electricity market clearance, and long term demand response. The latter addresses the complementary or substitutive nature of electricity consumption with other technologies according to price, as well as load shifting, and the consumers ability to modify load profiles to take advantage of persistently lower prices during periods of lower overall load.

Building on the same fundamental principles stated when addressing load, we develop an endogenous representation of reliability in Section 4.3. This can be incorporated into the wider framework and enables the parameterisation of the reliability of each technology and results in an equilibrium that is internally consistent with satisfying load requirements, reflecting both the imperfect reliability of additional capacity scheduled to buffer reliability issues, as well as the relevance of reliability issues of the marginal technology for the PDC. Our approach could be extended, and we discuss where it sits amongst the spectrum of possible approaches to modelling reliability.

Finally, we return to the foundation of the LDC, which is a chronological load pattern (CLP) as observed each day in electricity markets around the world. The broader focus is the incorporation of the impact of intermittent technologies, such as wind and solar generation, on the LDC in the model. To ensure the correct correlations are captured we use a daily chronological pattern and develop a set of complementarity conditions to transform between the LDC and the CLP. This enables realistic interpolation of pricing, which if done at an LDC level would not have been possible. We retain an adjustment function to capture the additional variability in the LDC relative to the fitted CLP.

## *4.2 Demand Response*

### 4.2.1 Introduction

The demand for electricity is multi-faceted. In the short-run, electricity spot markets face several demand-side challenges that make electricity markets susceptible to inefficient economic outcomes:

- Electricity prices are highly volatile relative to many other commodities or energy sources.
- Electricity spot prices are frequently updated and difficult to monitor.
- Electricity consumption is often difficult to adjust in response to changes in electricity spot prices.

- Electricity contracts often mute the incentive to adjust consumption according to electricity prices, even when that is possible.

Spot price volatility is the result of several factors and is exacerbated by significant demand-side correlations so that rather than many consumers, using electricity in an independent fashion, there are daily, seasonal and possibly annual patterns which are common to many participants. The fact that spot prices are produced as frequently as every five minutes requires that consumers must invest significant effort or expense if these prices are to be monitored. Even when prices can be monitored, electricity use is often part of a larger economic process, which cannot be easily halted or even adjusted. Nevertheless, there will be occasions when specific demand response opportunities, parameterised by an electricity price and facility cost, are available and we discuss these in Section 4.2.2.

As a direct consequence of the high price volatility, a lack of ability to monitor prices and a lack of ability to adjust consumption, there is also a significant degree of retail contracting. Among a wide spectrum of contracts available, some contractual forms preserve the possibility of price response by including price-based triggers for reducing load, while other contracts, such as those that may fall under the oversight of governmental or regulatory authorities that are favourable to consumers and may not even specify the maximum consumption available at the contract price. These contracts may completely negate any price based incentive for demand response. Depending on the overall mix of contractual forms and the level of uptake of each type of contract, consumers, in aggregate, may have negligible incentives to adjust their own demand in response to spot market prices.

The factors identified above lead to the standard and reasonable assumption that the spot market demand for electricity is inelastic in the short term. Our approach is slightly more nuanced, in that we assume the demand for electricity is inelastic in the short term, except for the various demand response technologies or initiatives specifically installed or designed for the purpose. Although the capability to respond to electricity prices is limited in the short term, over longer timeframes pricing patterns will emerge and will influence consumers planning electricity consumption, and in the plans that retailers offer. Our approach maintains the clear decision structure already established in our framework, namely that the investment and operational actions are distinct in time, and not simultaneous. Accordingly, in our framework, investors must invest in demand response technologies, and/or adjust consumption patterns in advance. As far as the latter is concerned we investigate two forms of equilibrium response to spot market prices; overall shifts in demand, perhaps to alternative fuels, and demand shifting to other periods throughout the season. Only then, having determined demand and the short term means of adjusting it, can the spot market be cleared.

### 4.2.2 Short Term Demand Response

With current technologies, the ability of demand-side participants to respond to pricing in the spot market typically requires prior planning and technological installations that are able to adapt in real time. Of paramount importance to demand response of any kind is the availability of actionable price information. The extent of the response that is desired depends on the circumstances of the consumer. However, irrespective of the desired response, the actual response of the consumer is largely influenced by the flexibility of the fixed stock of energy consumption devices being used and the method by which

actionable price information is discovered and translated into changes in energy consumption. Obtaining, actioning and ensuring enough flexibility exists to adapt represents an investment on the part of the consumer for which the returns must justify the outlay.

The load response options we envisage may literally be a technology or device, or it could be something as nebulous as a government initiated advertising campaign to reduce usage during a period of crisis. It is clearer in the former context that an actual demand response technology requires investment, but even the example of a government conservation campaign can be represented as a demand response option with a fixed and marginal cost. We can view such installations or programs as being analogous to a technology, with a cost of installation and a strike price, corresponding to the marginal cost component of a thermal technology, at which the demand response occurs, and is "in the money".

Once the investment decision has been made, the capacity of each demand response technology represents the fixed amount of demand that can be withdrawn when electricity prices reach a certain level, as defined by the marginal cost we associate with that technology. Once built, each demand response technology is treated as an individual technology with a fixed capacity in each sub-period. Accordingly, demand response technologies are subject to the same market clearance process as other generation technologies.

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} + \varphi^{+}_{i,r,t}\Big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i \in D, r, t \tag{4.1}$$

While the introduction of demand response options will alter the equilibrium plant mix, the value of the demand response technology in each sub-period is still defined by $\chi_{i,t}$ so that the optimal trade-offs between demand response technologies and generation technologies are also defined as before:

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC}_{\mathrm{j,t}} - \mathrm{MC}_{\mathrm{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i \in D, j \neq i, t \tag{4.2}$$

Where, in each sub-period, we have the following valuation of the demand response technology:

$$\chi_{i,t} - \sum_{r<R} \varphi^{+}_{i,r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i \in D, t \tag{4.3}$$

As with other technologies, there may be limits to the opportunities available. These might arise from production schedules, technical requirements, or safety issues, that limit the maximum quantum of the response associated with a particular opportunity.

$$CAP_i - CAP_i^{-} \geq 0 \quad \perp \quad \chi_i^{-} \geq 0 \qquad \forall i \in D > 0 \tag{4.4}$$

$$CAP_i^{+} - CAP_i \geq 0 \quad \perp \quad \chi_i^{+} \geq 0 \qquad \forall i \in D > 0 \tag{4.5}$$

The overall opportunity limit, $\mathrm{CAP}_{\mathrm{i}}^{+}$, relates to the level of available demand response for each type of demand response and is implicit in the nature and level of the load. This limit is determined by a broad range of factors, and so is "fixed" from an electricity sector planning perspective. Although somewhat redundant, for consistency we maintain the theoretical possibility of a minimum bound, so that, subject

to the bounds on demand response opportunities sites, the equilibrium capacity condition for each demand response technology is as for any other technology:

$$\mathrm{FC_i} - \sum_t \mathrm{w_t}\chi_{i,t} + \chi_i^+ - \chi_i^- \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \in D, t \qquad (4.6)$$

It may also be the case that the availability of demand response is not simply a capacity issue. Certain processes may be able to be abandoned for a length of time, but not indefinitely. In these cases, the use of the demand response opportunity must be husbanded to those periods when the savings from reducing consumption are greatest. In such a case, additional opportunities will exist at the margin, but will not be exercised because of an over-riding constraint that provide a floor for the aggregate consumption of the consumer. Section 3.4 detailed the formulation of technologies with both energy and capacity limits, and those formulations apply analogously here.

The underlying LDC remains unaltered as we treat notional (or actual) demand response technologies as generation technologies. Where economically viable, the inclusion of these technologies alters the set of critical utilisation levels and defines an intermediate price at which increases in load are absorbed by demand response technologies, up to the installed capability to do so. At the same time, the installed capacity of adjacent physical generation technologies reduces until equilibrium between the profitability and costs of those other technologies is restored. Notwithstanding the above, we can also net the load response off the underlying LDC to determine the observed and endogenous net LDC, and the level of generation capacity required from standard generation technologies.

$$L_{r,t}^{net} = L_{r,t} - \sum_{i \in D} GEN_{i,r,t} \qquad \forall r, t \qquad (4.7)$$

The PDC is defined as before, although it may include prices relating to the strike price of demand response opportunities.

Demand side bidding schemes generalise the demand response discussed above and offer the opportunity for demand side participants to shut down various processes in a graduated fashion in response to a range of different electricity prices. Just as a single consumer might have identified several critical electricity price points, each corresponding to a load response opportunity in the context of their individual economic circumstances, the market as a whole may present a continuum of such points. One approach to accommodating this extension is by adapting the configurable technology theory presented in Section 3.5. When viewed as a simple technology, the quadratic technology corresponds to a technology with a linearly decreasing marginal cost, as it becomes more efficient.

### 4.2.3 Long Term Demand Response

In the long run consumers can deduce pricing patterns and have far greater ability to adjust their consumption patterns. To the extent they can do so, that phenomenon represents a long term response, requiring both analysis and planning, which we discuss in this section. Changes in the general level of electricity prices can modify consumption by incentivising consumers to switch to, or away from, alternative fuels, or invest more or less in energy saving appliances. Although we do not deal with a general equilibrium model, electricity price levels naturally interact with other markets, causing, for

example, substitution between energy sources or adjustments in final good costings and output, which lead back to electricity demand. From a theoretical point of view, defining those relationships does not pose a problem, although in practice it requires development and calibration of a model that estimates the set of parameters most likely to generate the observed load pattern. For our purposes, it suffices to assume that a significant portion of those effects can be neatly summarised by own-price effects. Therefore we focus solely on consumption patterns and own-price effects on the demand for electricity and we decompose demand response into that which is due to movement in the overall level of electricity prices and that which is due to inter-temporal price effects, based on the relative price of electricity across different periods.

The response of load to the general level of prices can be characterised as a decision to reduce/increase consumption based on the bundle of prices corresponding to the consumers LWAP (Load Weighted Average Price). LWAP is particular to each user, or at least each distinct consumption pattern. We proceed assuming a single load profile although our approach could be generalised to include many load profiles. We also note that, primarily because of transaction costs, the contract market in most electricity markets provides for the consideration of a limited number of load profiles, and consumers must choose a less than perfect representation of their individual profile when contracting. Importantly, whether the long term demand response mechanism is through the contract market and based on load profiles built into contracts, or whether it is based simply on consumers observing the general level of prices, the resulting change in the load pattern will be identical, absent other differences which exist between spot and contract markets.

The second type of response we consider is the re-scheduling of consumption to less expensive periods or seasons. This is a response to relative prices by consumers who are exposed to multiple prices according to the period of use. These incentives are often expressed in a variety of contractual forms, ranging from household level incentives such as cheaper overnight rates, through to more complex contracts for industrial and manufacturing users, with TOU (Time of Use) pricing structures designed to incentivise the orientation of production towards less expensive periods of electricity generation. Unfortunately, electricity demand is often not inter-temporally transferrable as, for example, households are often unable to shift heating and/or cooling needs to other times of the day. Similarly, industrial users have significant scheduling, operational and institutional issues that restrict them from adjusting consumption even over longer timeframes. Nevertheless, the fact that such contractual options do exist, and have persisted, suggests that consumers do have some, albeit limited, ability to shift their load.

Over longer timeframes, the potential for both types of demand response are greatly increased, and this demand response is by nature an adjustment of underlying load, which we have assumed to be exogenous until now and which we define without loss of generality to be the load level that would occur if the price of electricity were zero. The possibility of long term demand response through changes in consumption patterns affects the viability and equilibrium role of other technologies, including demand response technologies as discussed in Section 4.2.2, for which longer term load adjustment is a direct substitute.

For illustrative purposes, we introduce a linear demand function where overall adjustment of the base level of demand $L\left(\lambda_r, \lambda^{avg}\right)$ is in response to average prices, $\lambda^{avg}$, while demand shifting occurs linearly according to relative prices, $\lambda_r - \lambda^{avg}$. The coefficients $\mathrm{a}^{\mathrm{shift}} > 0$ and $\mathrm{a}^{\mathrm{avg}} > 0$ respectively determine the strength of these effects. The functional form need not be linear and other forms, such as a constant elasticity form, may have desirable properties that a researcher may wish to pursue. The demand shifting capability in this case is defined within a season or sub-period, so that load will transfer from high price periods within a season to lower price periods within that season. That might reflect shifts from high price periods in a day, to lower price periods in a day, but might also reflect a shift from one week to another, for example. Where load shifting is contemplated, the adjustment can equally be applied to a chronological load pattern as shown in Section 4.4.2, enabling the implications of demand shifting on daily operations to be better reflected.

The demand function has the following form:

$$\begin{aligned} L_{r,t} &= \mathrm{L}\left(\lambda_{r,t}, \lambda^{avg}\right) \\ &= \mathrm{L}^{0}_{r,t} - \mathrm{a}^{\mathrm{shift}}\left(\lambda_{r,t} - \lambda^{avg}\right) - \mathrm{a}^{\mathrm{avg}}\lambda^{avg} \end{aligned} \qquad \forall r \qquad (4.8)$$

Analogously, we could be analogously discussing the price response in terms of contract prices in a risk neutral setting. Therefore, whether consumers are reacting to contract prices or directly to general price levels, we can use the following definition of load weighted average pricing to represent general pricing effects:

$$\lambda^{avg} = \frac{\sum_{r,t} \lambda_{r,t} L_{r,t}}{\sum_{r,t} L_{r,t}} \qquad (4.9)$$

Substituting the definition of average prices and generalising the demand response parameters to a sub-period level, perhaps to reflect varying degrees of demand elasticity, we have:

$$L_{r,t} = \mathrm{L}^{0}_{r,t} - \mathrm{a}^{\mathrm{shift}}_{t}\left(\lambda_{r,t} - \frac{\sum_{r,t} \lambda_{r,t} L_{r,t}}{\sum_{r,t} L_{r,t}}\right) - \mathrm{a}^{\mathrm{avg}}_{t}\frac{\sum_{r,t} \lambda_{r,t} L_{r,t}}{\sum_{r,t} L_{r,t}} \qquad \forall r,t \qquad (4.10)$$

Each price change induces two effects; a global effect based on the impact of average or general price levels, and a substitution effect based on the shifting of load away from high price periods to low price periods. For example, an increase in the equilibrium spot market price $\lambda_{r,t}$ results in a demand shift from the period corresponding to r away to other relatively cheaper periods. There is also an overall reduction in load on account of the general increase in prices. In other periods, the increase in $\lambda_{r,t}$ has an ambiguous effect. Load shifts to other periods may or may not dominate the overall load reducing effect of higher prices. Working backwards from the observed LDC, if we unwind both the overall and substitution effect, the underlying LDC must not only reflect higher load, but also peakier load.

Manipulating (4.10), we have:

$$L_{r,t} = \mathrm{L}^{0}_{r,t} - \mathrm{a}^{\text{shift}}_{\text{t}}\lambda_{r,t} - \left(\mathrm{a}^{\text{cont}}_{\text{t}} - \mathrm{a}^{\text{shift}}_{\text{t}}\right)\frac{\sum_{r,t}\lambda_{r,t}L_{r,t}}{\sum_{r,t}L_{r,t}} \qquad \forall r,t \qquad (4.11)$$

Given a specification of the relationship between market prices and load, the most common approach to modelling demand response is to substitute the definition into the market clearing constraint. In our framework, this approach yields the following market clearing constraint:

$$\sum_{i} GEN_{i,r,t} - \left( \mathrm{L}^{0}_{r,t} - \mathrm{a}^{\text{shift}}_{\text{t}}\lambda_{r,t} - \left(\mathrm{a}^{\text{cont}}_{\text{t}} - \mathrm{a}^{\text{shift}}_{\text{t}}\right)\frac{\sum_{r,t}\lambda_{r,t}L_{r,t}}{\sum_{r,t}L_{r,t}} \right) = 0 \;\perp\; \lambda_{r,t}\ free \quad \forall r,t \qquad (4.12)$$

The implication of this transformed complementarity condition is that as the market price increases the level of load required to be serviced decreases. It may or may not be the case that this load response is marginal and therefore price setting, but it is certainly a possibility. It is possible that the modeller intends to model a near instantaneous price response in electricity markets, and in the case where that response can occur in the timeframe of the market clearance, without requiring investment. In this case, then the modeller is justified in introducing the definition of load into the market clearing constraint shown in (4.12). Unfortunately, for load response of the same genesis as we have discussed here, this approach is inconsistent: long-term demand responses cannot directly influence market clearing prices. The inconsistency mirrors the inconsistency identified in Chapter 1.

Ideally, we would like to incorporate this load response by treating it as a technology. This preserves the analogy of the previous section. The fundamental difference though, is this is an unresponsive technology in the short term, and therefore we must define the load response in such a fashion that the capacity of the response is registered against the underlying load without being price sensitive in the short term. Accordingly, we define a new notional technology with capacity equal to the physical quantity of load response, but with a marginal cost of zero, so that in the absence of negative prices, it is always included and never marginal. The capacity of the technology is defined as follows:

$$CAP_{LR,r,t} - \mathrm{a}^{\text{shift}}_{\text{t}}\lambda_{r,t} - \left(\mathrm{a}^{\text{cont}}_{\text{t}} - \mathrm{a}^{\text{shift}}_{\text{t}}\right)\frac{\sum_{r,t}\lambda_{r,t}L_{r,t}}{\sum_{r,t}L_{r,t}} = 0 \qquad \forall r,t \qquad (4.13)$$

Where $\mathrm{MC}_{\mathrm{LR,r,t}} = 0$. The model now includes I conventional technologies along with a notional shortage technology, for which i=0, and a notional load response technology, which we denote as LR, but index with i=I+1. The market equilibrium conditions remain unchanged barring the extension to the technology index. The market clearing constraint reflects the reduction in load as a result of price by the inclusion of the notional load response technology in all market clearances on account of its marginal cost.

Unlike other technologies the capacity of load response is not fixed as it is determined endogenously. Accordingly, it is unnecessary to constrain the capacity of load response to some

consistent value as is the case with conventional technologies. It is also not subject to investment cost, within the context of the problem.

The inclusion of load response will naturally reduce investment in other generation technologies. Of particular interest is that long term load response is a substitute for short term load response. While the two approaches to load response deliver consumers similar benefits, they are not identical. Neither are they from the perspective of the market. Long term load response is directed at those technologies involved in supply of all of the consumer's load profile, whereas the short term technological response is directed primarily at high price periods and so represents a relatively direct alternative to peaking technologies.

## *4.3 Plant Reliability*

### 4.3.1 Calculating Reliability

We begin by clarifying the nature of reliability that we investigate. At one end of the spectrum, we are not considering the response of the system to breakdowns. Those outages will be compensated for in an economically efficient fashion with contingency services. When a plant breaks down, the extent to which a breakdown requires response from the market is somewhat dependent on the nature of the offer process or whether there is a day-ahead market operating in which the plant has been offered for generation. Beyond the immediate periods following a breakdown contingency responses no longer apply and additional capacity from some other source is required to satisfy load. At the other end of the spectrum, we are not considering plant outages that persist for multiple periods, such as might occur for reasons of safety after a nuclear, or other, accident. Reliability issues of this sort are better assessed in an explicit stochastic framework. Instead of these polar issues, we consider reliability issues such as planned shut downs for maintenance where it is known in advance that the capacity will not be available.

A comprehensive approach to reliability, or any other insertion of variability into the load duration curve, properly requires a full convolution of the distribution of reliability with the distribution of load. Such a convolution would, in theory, contemplate both the limiting cases; when all capacity was available, along with the case when no capacity was available. An intermediate approach might assign reliability factors at the plant level and, based on typical plant sizes, create a binomial approximation of the reliability distribution which through convolution could be incorporated in the LDC.

For each technology i, we define its reliability as $\rho_i$, so that $1-\rho_i$ is the proportion of time the capacity of technology i is unavailable for generation. Our approach stops short of convolution, but does address the endogeneity of the reliability by calculating the expected reliability of the plant mix at each utilisation level based on the composition of the generation mix at each utilisation level. In support of our approach, we note that the LDC will typically dominate any convolution with the reliability distribution. In any case, we adopt a point estimate of the capacity of technology i available for generation. That estimate is equal to the expected available capacity of technology i, with installed

capacity $CAP_i$, and is given by $\rho_i CAP_i$, so that expected quantity of capacity of technology I unavailable for generation is given by $(1-\rho_i)CAP_i$.

We do not concern ourselves with impaired capacity, in which the plant is operating but its efficiency is reduced. Instead, the quoted availability of the plant, $\rho_i$, is an assessment of reliability that represents the equivalent availability of the full plant, appropriately factoring both full breakdowns and partial impairment situations. We also do not consider the optimisation of scheduled outages on the basis of pricing. Those wishing to formalise this aspect of operations could explore an explicit model of outages or the adjustment of the reliability of each technology to reflect the sub-periods in which scheduled outages are most likely.

In the spirit of Baldick (2009), we propose augmenting the LDC by the expected outages given the generation mix for the corresponding load level, and allowing the standard market clearance procedures to define the response to outages. In adopting this approach, we treat all existing capacity as reliable for the purpose of the market clearing process and augment load, rather than reduce the available capacity of each technology to account for outages. To do so, we need to identify the average level of material, or relevant, outages, as distinct from the expected level of outages for a given level of capacity. Where a technology is supramarginal, its availability or otherwise is not relevant. With respect to that technology, relevant outages are zero. This contrasts with the case where of infra-marginal technologies, where all outages are relevant, and alternative generation from higher marginal cost technologies is required to satisfy load.

The relevance of outages of the marginal technology is dependent on the balance between the level of generation required of that technology and the level of outages experienced by that technology. To define material outages as a function of generation, we could scale expected outages by $GEN_i / CAP_i$ so that they are defined as $(1-\rho_i)GEN_i$, in terms of generation, but while this scaling correctly assesses the relevance of outages in the polar cases of infra-marginal and supra-marginal technologies, it is not the case that the relevant outage level for the marginal technology is a proportion of generation. The scaling approach implies that even when a small fraction of available capacity is required for generation the unavailability of certain plants of technology i proportionally impacts the aggregate generation capabilities of technology i. But when faced with a schedule of unavailable plants, operators will not schedule those plants. Therefore, those outages are not relevant until that capacity is actually required. When $0 < GEN_{i,r,t} \leq \rho_i CAP_i$, then technology i is marginal and there is sufficient unused capacity available to compensate for outages of technology i, so that those outages are able to be accommodated and are not relevant. Conversely, when $GEN_{i,r,t} > \rho_i CAP_i$ outages prevent technology i from servicing any additional load. We remind the reader that while this scenario appears to be infeasible it is not, as we have elected to augment load rather than penalise capacity in our formulation.

We define expected outages, $OUT_{i,r,t}$, of technology i as:

$$OUT_{i,r,t} + \rho_i CAP_i - GEN_{i,r,t} \geq 0 \quad \perp \quad OUT_{i,r,t} \geq 0 \quad \forall i,r,t \tag{4.14}$$

When technology is supramarginal or marginal with comparatively low generation, then $GEN_{i,r,t} < \rho_i CAP_i$, so that the complementarity condition (4.14) requires that $OUT_{i,r,t} = 0$. Conversely, when technology i is inframarginal or marginal with comparatively high generation, then $GEN_{i,r,t} > \rho_i CAP_i$, the complementarity condition (4.14) records outages correctly as $OUT_{i,r,t} = GEN_{i,r,t} - \rho_i CAP_i > 0$.

At a given load level we define total relevant outages, $OUT_{r,t}$ by summing the outages of individual generation technologies as defined above:

$$OUT_{r,t} = \sum_i OUT_{i,r,t} \qquad \forall r,t \qquad (4.15)$$

### 4.3.2 Market Clearance

To define an equivalent perfectly reliable market clearance we augment load to account for outages. The market clearing condition becomes:

$$\sum_i GEN_{i,r,t} - \left(L_{r,t} + OUT_{r,t}\right) = 0 \quad \perp \quad \lambda_{r,t} \; free \qquad \forall r,t \qquad (4.16)$$

This approach recognises that the generation required to service the augmented load also must be included in the calculation of outages. In doing so, we recognise the possibility that capacity being used to cover the breakdowns of other plants is itself less than perfectly reliable. No restriction is placed on additional capacity of a technology compensating for outages of the same technology.

The other market clearing conditions require no further adjustment. Neither do the optimal trade-off definitions, although the actual optimal trade-offs will adjust according to the implications of reliability on the capacity of each technology.

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^e_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^e_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \qquad (4.17)$$

### 4.3.3 Investment

The expected direct costs associated with any breakdown or maintenance are assumed to be included in the fixed operating costs for each technology, perhaps representing a contract price for maintenance or the cost of a life time guarantee, and these are amortised over the operational lifetime of the plant, leaving the investment implications of plant breakdowns limited to three areas.

The first implication of reliability is in the spot market itself. Each technology benefits from the unreliability of others. These benefits accrue naturally in our framework, through changes in market clearances. The adjustment of market clearances is driven by (4.16), which respectively requires the satisfaction of a greater "equivalent" load, and (4.17), which reflects the change in the imputed optimal trade-off in each sub-period arising from that. The sub-period earnings condition implicitly incorporates the implications of less than perfect reliability in the spot market:

$$\chi_{i,t} - \sum_{r<R} \varphi^+_{i,r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t} \; free \quad \forall i > 0, t \qquad (4.18)$$

Secondly, we must account for the outages of each plant from the perspective of its owner. In (4.18), the imputed value of capacity of technology i in sub-period t is assessed on the basis that the technology is perfectly reliable, in accordance with the augmentation of the LDC to account for outages. Therefore, the investor, while observing the PDC and the returns available, will note that the earnings of each technology must be scaled by its reliability factor as those market-based earnings will be forgone for the proportion of time that the technology is unavailable.

$$\mathrm{FC_i} - \rho_i \sum_t \mathrm{w_t} \chi_{i,t} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{4.19}$$

The scaling in (4.19) could be developed significantly if the reliability of the plant could be demonstrated to be seasonal, or perhaps related to heat or cold. It could further be used as a device for recognising the nature of planned or regular schedulable maintenance, by setting the factor to an appropriate fraction in those seasons it is most advantageous to shut down and carry out maintenance.

An alternative, but equivalent, interpretation of (4.19) is to cast the adjustment in terms of costs. By adjusting fixed costs, we can reflect the imputed cost of supplying a reliable unit of capacity that could achieve the earnings defined in (4.18) that are available to perfectly reliable technologies:

$$\frac{\mathrm{FC_i}}{\rho_i} - \sum_t \mathrm{w_t} \chi_{i,t} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{4.20}$$

Depending on the interpretation chosen, the greater the reliability of a technology, the less the adjustment to fixed costs, or the discount to earnings.

As the reliability of a technology worsens, its involvement in the equilibrium plant mix will reduce. Relative to the counterfactual of perfect reliability, global optimal trade-offs tend to swing in favour of low capital intensity technologies as less than perfect reliability represents a tax, or additional rate of return requirement, on the fixed costs of each technology. However, it is not necessarily the case that an unreliable technology will fare worse than it does in the perfectly reliable counterfactual. Given equal rates of reliability, the burden of the adjustment is unambiguously on more capital-intensive technologies. In the limit, a relatively reliable baseload technology could be eliminated from the equilibrium plant mix, while a relatively unreliable peaking technology might expand its utilisation. When viewed in terms of cost adjustment, the underlying reason is that the fuel or operating component of the cost structure is fully "reliable", whereas the capacity component is not. The former comprises a significant portion of the peaking technology's cost structure, while it is a smaller portion of the total cost of a baseload technology, making the issue of reliability more important to high capital cost technologies.

Unsurprisingly, it is typically the case that more capital intensive technologies are more reliable than low capital intensity technologies. Apart from the operating characteristics being suggestive of more consistent, and presumably beneficial, usage patterns, if we consider the case of a single technology, the reliability factor could be optimised based on the trade-off between different configurations or installation options, presumably at a cost. By the above logic, such an adjustment would be more valuable for more capital intensive technologies, ceteris paribus, and so we should expect to see highly capital intensive technologies installed with higher optimised reliability levels.

The change in the equilibrium role of a single technology is dependent on the characteristics of all other technologies and is therefore somewhat ambiguous. The exceptions are the notional shortage technology and, if it is included, the notional demand response technology, which we can assume to be perfectly "reliable". When viewed from a technological perspective, the implication of recognising the less than perfect reliability of conventional technologies is unambiguous; the shortage frequency will increase as the notional shortage technology is both perfect reliability and has zero fixed costs.

Finally, we have assumed the duration of outages is relatively short. However, it may be the case that certain outages are significant and could spread beyond the sub-period structure of the model. We do not address these prolonged outages but note the scope of response depends significantly on the timeframe of the outage so that, in a system with specific concerns in this respect, other decision-making processes would be influenced whenever prolonged outages were anticipated. In the limit, prolonged outages, such as those associated with the Japanese tsunami of 2011 can remove plants, or even entire technologies, from availability for years or possibly even permanently, thereby requiring further assessment of not only contract positions, but also investment plans. Where the nature of the reliability issue is such, our approach is inappropriate as it assumes the reliability of a technology can be treated deterministically as a proportion, in precisely the same way the framework treats load in less than a fully stochastic basis. In those cases, scenario development would be preferable, as would the likely development of risk measures surrounding those possibilities.

## *4.4 Intermittent Generation*

### 4.4.1 Introduction

So far we have discussed technologies that are dispatchable but, in the case of intermittent generators, the energy that enables generation of electricity arrives according to a random natural process. Without supporting storage, the critical distinguishing feature of these technologies from the perspective of the market and investors, is the relative lack of control over generation. The most prominent intermittent technologies are wind, solar, and run of river hydro, and the integration of these technologies into electricity markets poses challenges of a technical and economic, nature (Macgill, 2010). Investors in these technology and site combinations are focussed on selecting sites and technologies whose generation is highly correlated with higher price periods. When this is the case, the random process that drives generation mimics the sort of intertemporal allocations they might make if storage was available.

A thorough approach for calculating the net load duration curve would take account of the potential correlation between load and intermittent generation, as well as the variability of intermittent generation which depends greatly on the correlation between different intermittent producers. To achieve this fully requires a full convolution, but this computationally challenging in light of the rest of the structure and given the individual characteristics of different technologies.

We could consider formulation of a model in terms of the correlation between the generation of each intermittent technology and load. But a single measure cannot satisfactorily portray the underlying relationships. In any case, the relevant correlation is not the correlation between the output

of a technology and load, but between the output of the intermittent technology and net load, which if we are to continue the finance analogy represents the performance of the rest of the market. While this definition could be accommodated, a larger problem exists. Correlation itself can be poorly estimated and understood where the functional form of the relationship is poorly specified or not accurately characterised. For example, it is well known that a relationship that is precisely quadratic in nature can produce zero correlation when the correlation is assessed on a linear basis. Ideally, any approach to modelling intermittent generation should address the underlying features of the system such as the daily patterns that typically drive the correlation between load and the output of individual technologies, rather than rely on a summary measure of correlation that is clouded by other effects.

Our analysis of intermittent generation focusses on the correct representation of the correlation between intermittent generation and net load on an intra-day basis. We do not consider the full distribution of intermittent generation, just its pattern. The reasonableness of this approach rests on the size of load, the variation of which is included in the LDC, relative to the size of variations around the average output of intermittent generation at each time of day. It follows that as market penetration of intermittent generation increases, the requirement to address the full distribution of intermittent generation does also, and the output of the model we propose should be subjected to post-solution scrutiny to either ensure the approximation is suitably accurate, or possibly to investigate an endogenous version of the adjustment function.

To capture the chronological correlations that exist, we adapt our framework to address the modelling of intermittent generation technologies that exhibit strong intra-day generation patterns. We begin by determining chronological generation functions and a chronological load pattern to enable more detailed assessment of the relationship between generation and load, and other generation technologies if required. In doing so we deconstruct the LDC into two components; an LDC equivalent to the chronological load pattern, which itself is endogenous and an adjustment function that enables translation between each load representation. As investment in intermittent technologies occurs, the net chronological load function is formed and converted to an estimated net LDC that is serviced by conventional generation technologies

### 4.4.2 Chronological Load and Generation

Because the LDC is a convolution of various load determinants, a given load level might correspond to a cold morning, or a warm afternoon, or some other combination of load determinants. That combination of influences obscures the relationship between intermittent generation and load. As the seasonal granularity is increased, the variation within the LDC that is associated with seasonal effects is reduced, so that the proportion of variation that is associated with daily variations increases. Increasing the model granularity successively removes an increasing quantity of seasonal variation, so that in the limit the resulting LDC contains only daily variations. To clarify the nature of correlation between intermittent generation and load, or net load, we need a chronological representation of each. To that end, we introduce a representative chronological load pattern (CLP), which describes the pattern of load throughout the day, and can be defined either globally or on a sub-period or seasonal basis. A similar intermittent generation pattern can be developed for each individual intermittent

technology so that in combination with the CLP, the actual relationship between load and individual intermittent generation technologies is defined.

As shown in Figure 27 a representative, or average, piecewise linear chronological load and generation pattern can be formed.

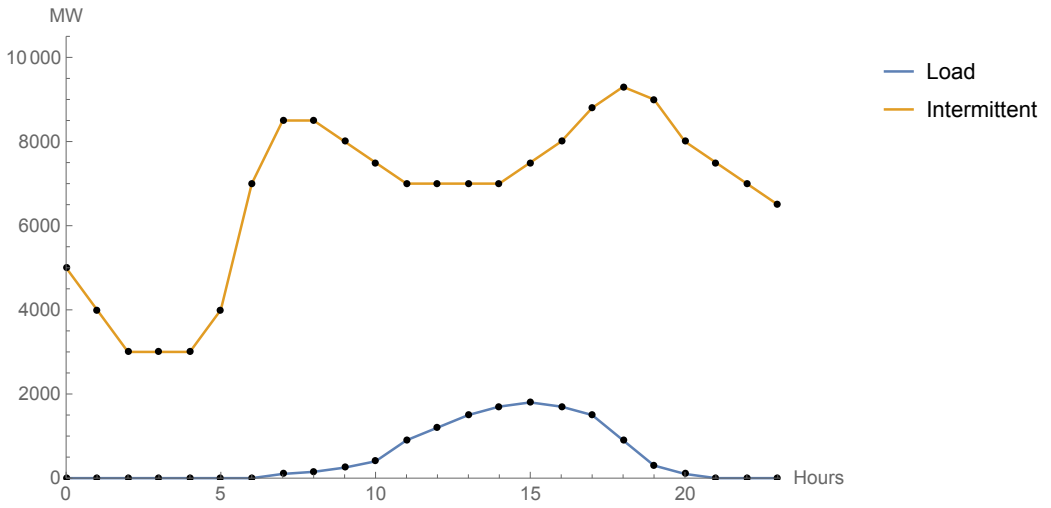



**Figure 27: Chronological Load Pattern & Intermittent Generation**

As with thermal technologies, some judgement is necessary when grouping individual technologies. While two installations of the same underlying technology may be ostensibly similar, they may vary significantly in performance. Intermittent technologies require more acute attention to be given to this issue, as intermittent technologies are not installed in a homogenous fashion, as their output typically depends significantly on location. Accordingly, to preserve any differences in the typical generation profile, we would ideally maintain the distinction between different installations of the same technology, as these should only be aggregated only when individual installations are highly correlated with each other, as assessed on an intra-day basis. Naturally that ideal must be balanced with computational feasibility.

Unlike the LDC, the CLP does not represent the full distribution of load levels occurring within a season, so to ensure the most accurate representation we utilise a CLP in conjunction with the LDC. The CLP represents the average, or best fit, chronological load. Where the intra-day pattern is most relevant, the average load profile as expressed by the CLP provides a useful basis for analysis of daily correlations, but where aggregate load and energy levels are more relevant, we reference an LDC.

Genuinely different CLP's, such as for weekdays and weekends, might exist in some markets, and, as CLP's are not intertemporally linked in this framework, our approach is readily generalised to accommodate more than one CLP when distinct load patterns are deemed significant, even when those CLPs apply to non-contiguous time periods.

### 4.4.3 Formulation

*Decomposing the LDC*

From a CLP, an equivalent LDC can be formed. We normally transform the initial chronological load pattern into an LDC. Conceptually we are re-ordering the CLP to get an LDC consistent with that average chronological pattern. The dynamic approach detailed below could also be invoked to perform this task although it is rather cumbersome for a one-off calculation when the data required to seed the problem is exogenous.

Whereas the LDC implicitly incorporates all observations, the CLP, estimated using those same observations is, depending on the chosen estimation criteria, a "best-fit" with estimation errors. Even setting aside issues of estimation, the CLP equivalent LDC, $L^{CLP}$, is inconsistent with the actual LDC, from a purely conceptual perspective. Unless all of the intra-season variation in load is present in the chronological pattern represented in the CLP, the LDC will be peakier than $L^{CLP}$, as shown in Figure 28. For given utilisation levels, the load levels implied by the chronological pattern will not precisely coincide with the LDC, which represents the full distribution of load levels throughout a season. The difference will be due to variability not explained by the estimation of the CLP.

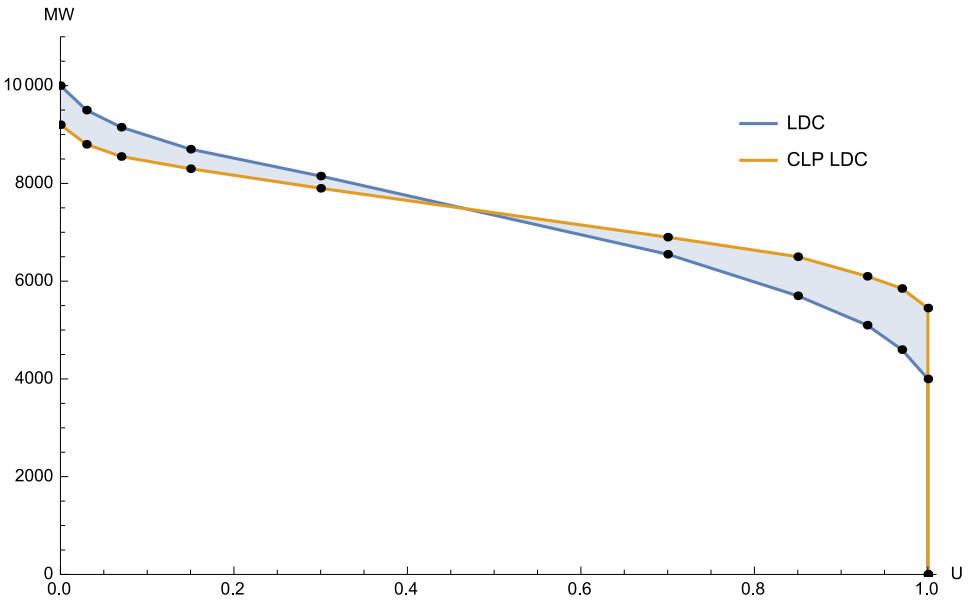



**Figure 28: Original LDC vs. CLP Equivalent LDC**

To at least preserve the variability incorporated into the LDC and enable conversion between the two load descriptions, we define an adjustment function that describes the difference between the two load measures in a form that may be re-applied to adjusted load patterns within the model. To guarantee consistency between a piecewise linear CLP equivalent LDC and the original LDC, the adjustment function must also be piecewise linear, with interpolation occurring between the complete set of K+1 utilisation levels and H+1 intra-day markers that respectively define the LDC and the CLP. As shown in Figure 28, these sets will generally not have common elements. In addition, the load levels associated with the chosen utilisation levels and intra-day markers will also not correspond with one another.

Using l=1… L as an index of the combined set of points used to define the LDC and CLP equivalent LDC. Adopting a proportional scaling method rather than an absolute difference, we define the scale factor at $u_l$ as:

$$\text{SCALE}_l = \frac{L_l^0}{L_l^{CLP}} \qquad \forall h \qquad (4.21)$$

Here $L_l$ is the load level on the piecewise linear LDC corresponding to $u_l$ and $L_l^{CLP}$ is the load level corresponding to $u_l$ on the piecewise linear LDC formed from the chronological load pattern. We define the scale factor at other utilisation levels in similar fashion.

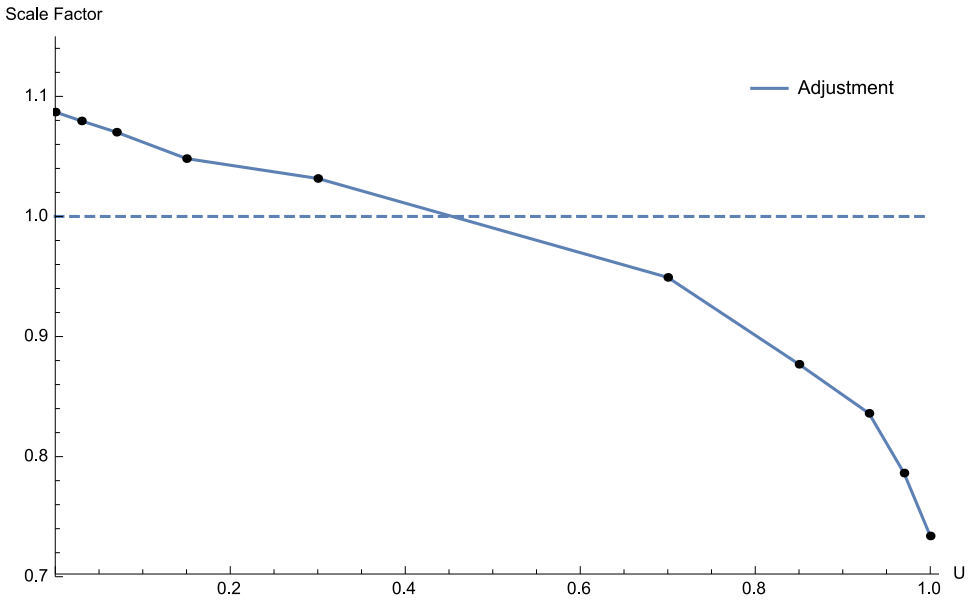



**Figure 29: Adjustment Function**

As shown by way of example in Figure 29, the adjustment function is not consistently greater than or less than unity, although it should tend to be decreasing in utilisation as the LDC represents a wider distribution than a set of averages based on unordered load levels would, and so is more extreme in the tails of the load distribution, which correspond to zero and full utilisation. As some of the utilisation levels in the model are determined endogenously, we seek a mathematical representation of the adjustment function shown in Figure 29 so that we can calculate a scale factor and apply it to an endogenously calculated CLP equivalent LDC.

One approach is to estimate a function that describes the scale factor as a function of utilisation levels as in (4.22) to (4.25). For the same reason as above, this function must also be monotonically decreasing. Several functional forms could be used but for illustrative purposes we propose a cubic function, as it can reasonably approximate the characteristics of scale factor divergence at the extremities of the utilisation range.

$$\underset{a,b,c,d}{Minimise} \sum_l \left(S(u_l) - \text{SCALE}_l\right)^2 \qquad (4.22)$$

$$S(u_l) - \left(a + bu_l + cu_l^2 + du_l^3\right) = 0 \qquad \forall l \qquad (4.23)$$

$$b + 2cu_l + 3du_l^2 \le 0 \qquad \forall l \qquad (4.24)$$

$$a, b, c, d \text{ free} \qquad (4.25)$$

In the above optimisation the objective is rather arbitrarily defined as the sum of the squared errors from estimation, that being a relatively standard criteria. The estimation function is a general cubic form and the optimisation is over the coefficients of the function. The gradient is restricted to be weakly negative at all utilisation points in the set L. This approach does admit the possibility that the function could exhibit a positive gradient between two points being interpolated so it may not be strictly monotonically decreasing at all utilisation levels. In addition, the choice of a non-linear scale factor does suggest a potential inconsistency with the rest of the development, which is generally piecewise linear.

A second approach is to define a piecewise linear function, interpolating between the points as shown in Figure 29, and using an approach similar to Section 2.6, define the adjustment factor according to the utilisation range of interest. To ensure this approach works, we must ensure that the scale factor is decreasing in utilisation so that the process of reconstruction does not lead to a non-monotonic, or partially ordered, LDC. A simple but ad hoc solution to the problem of ensuring monotonicity involves combining any segments where the behaviour is not monotonic until the average behaviour is. This is also a rather imprecise approach, and could be replaced with a dynamic determination of whether this is even necessary, as it is not the monotonicity of the adjustment function itself that is required, it is the monotonicity of the combined adjustment function and CLP consistent LDC that is relevant. The choice of the functional form of the adjustment function, is not critical, particularly when the underlying load is not endogenous as a result of demand response, for example. For the purpose of our discussion we designate that function S(u), without specifying its exact nature.

We have decomposed the underlying LDC into an alternative LDC, consistent with the CLP, and a set of scale factors. From those scale factors we have then developed an adjustment function, S(u), that approximates the scale factor as a function of the utilisation level. This procedure can be used once or, for example, in the case where we also wish to combine analysis of intermittent generation with price response, the estimation procedure that defines S(u) can be endogenous and rather than be calculated a priori, it can be included using the complementarity conditions corresponding to the optimisation that defines it.

*Determining the Net CLP*

To determine the net chronological load pattern we must first aggregate the output of intermittent technologies. There will be many such technologies, each with a different degree of correlation to load and different endogenous capacity, which together form a portfolio of intermittent generation that we must deduct from load. Our focus is on the intra-day relationship so we perform this task with the CLP, and leave seasonal variations to be addressed at the sub-period timeframe. We define the generation pattern by the capacity factor of each intermittent technology at each time of day,

$ICF_{i,h}\ \forall i \in INT$. Given a capacity of $CAP_i$, the total average generation at time h is given by $GEN_h^{INT}$ :

$$GEN_h^{INT} = \sum_{i \in INT} \mathrm{ICF}_{\mathrm{i,h}} CAP_i \qquad \forall h \qquad (4.26)$$

Accordingly, where $\mathrm{L}_{\mathrm{h}}^{\mathrm{CLP}}$ is the chronological load measurement at h, then $NL_h^{CLP}$, the net load at a particular h, is equal to the chronological load less the expected generation of intermittent technologies:

$$NL_h^{CLP} = \mathrm{L}_{\mathrm{h}}^{\mathrm{CLP}} - \sum_{i \in INT} \mathrm{ICF}_{\mathrm{i,h}} CAP_i \qquad \forall h \qquad (4.27)$$

*Re-constructing the Net LDC*

Whereas plant outages augment the load duration curve monotonically, the same is not true for intermittent generation. As a result of intra-day generation patterns, intermittent generation can alter the rank correlation between the time of day and the corresponding net load level. Therefore, applying the definition in (4.27) with the same ordering as for $\mathrm{L}_{\mathrm{h}}^{\mathrm{CLP}}$ does not suffice and we must dynamically construct a monotonic net LDC from the net chronological pattern and the adjustment function S(u) defined earlier.

To define a net LDC consistent with any given net CLP, we identify the net load levels, $NL_h$, and intra-day times, $t_h$, that define the CLP, where $t_h$ is expressed as the fraction of the day that has elapsed. Taking each net load level, $NL_h$, as a reference level, we then determine the proportion of that section of the net CLP that is above or below that load level. There are six possibilities, comprised of the permutations of net load increasing/decreasing with whether net load is always above, always below, or intersecting the reference net load level we are concerned with.

We begin by considering the cases in which the reference net load does intersect the net CLP within a particular time segment. When net load is increasing over the segment defined $t_{h-1}$ and $t_h$ then, using $h^* \in H$ as an alias for h to index reference net load levels, we have the following expression that defines the utilisation factor, $u_{h^*,h}^{CLP}$, corresponding to the reference net load level $NL_{h^*}^{CLP}$ such that $NL_{h-1} \le NL_{h^*}^{CLP} \le NL_h$ :

$$\begin{aligned} u_{h^*,h}^{CLP} &= \frac{t_h - t_{h-1}}{NL_h - NL_{h-1}} \left( NL_h - NL_{h^*}^{CLP} \right) \\ &= \frac{NL_h - NL_{h^*}^{CLP}}{NL_h - NL_{h-1}} \left( t_h - t_{h-1} \right) \end{aligned} \qquad \forall h, h^* > 0 \qquad (4.28)$$

A similar expression is available when net load is decreasing over the segment defined $t_{h-1}$ and $t_h$. The following expression defines the utilisation, $u_{h^*,h}^{CLP}$ within that segment for a reference net load level $NL_{h^*}^{CLP}$ such that $NL_h \le NL_{h^*}^{CLP} \le NL_{h-1}$ :

$$u^{CLP}_{h^*,h} = \frac{t_h - t_{h-1}}{NL_{h-1} - NL_h}\left(NL_{h-1} - NL^{CLP}_{h^*}\right) = \frac{NL_{h-1} - NL^{CLP}_{h^*}}{NL_{h-1} - NL_h}\left(t_h - t_{h-1}\right) \qquad \forall h, h^* > 0 \tag{4.29}$$

To combine expressions (4.28) and (4.29), we introduce two complementarity conditions that determine whether or not a particular segment exhibits increasing or decreasing net load. These conditions enable the selection of the correct numerator and denominator in the ratio of relative net load in the expressions (4.28) and (4.29).

$$NL_h - NL_{h-1} + \gamma^1_h \geq 0 \quad \perp \quad \gamma^1_h \geq 0 \qquad \forall h > 0 \tag{4.30}$$

$$NL_{h-1} - NL_h + \gamma^2_h \geq 0 \quad \perp \quad \gamma^2_h \geq 0 \qquad \forall h > 0 \tag{4.31}$$

When net load is increasing $NL_h > NL_{h-1}$ so from (4.30), $\gamma^1_h = 0$ and $\gamma^2_h = L_h - L_{h-1}$. Similarly, when net load is decreasing we have $NL_{h-1} > NL_h$ and $\gamma^1_h = L_{h-1} - L_h$, and $\gamma^2_h = 0$. Accordingly, the denominator in the above expressions can be recorded as $\gamma^1_h + \gamma^2_h$, giving (4.28) when net load is increasing and (4.29) when net load is decreasing. We can also define the numerator using the definitions of $\gamma^1_h$ and $\gamma^2_h$. When net load is increasing, we require $NL_h - NL^{CLP}_{h^*}$ and when net load is decreasing we require $NL_{h-1} - NL^{CLP}_{h^*}$. We define the numerator as follows:

$$\left(1 - \frac{\gamma^1_h}{NL_{h-1} - NL_h}\right)NL_h + \left(1 - \frac{\gamma^2_h}{NL_h - NL_{h-1}}\right)NL_{h-1} - NL^{CLP}_{h^*} \qquad \forall h, h^* > 0 \tag{4.32}$$

Combining (4.32) and the denominator, $\gamma^1_h + \gamma^2_h$, we have a general expression for $u^{CLP}_{h^*,h}$:

$$u^{CLP}_{h^*,h} = \frac{\left(\left(1 - \frac{\gamma^1_h}{NL_{h-1} - NL_h}\right)NL_n + \left(1 - \frac{\gamma^2_h}{NL_h - NL_{h-1}}\right)NL_{h-1} - NL^{CLP}_{h^*}\right)}{\gamma^1_h + \gamma^2_h}\left(t_h - t_{h-1}\right) \ \forall h, h^* > 0 \tag{4.33}$$

To verify this expression, we consider the example where net load is decreasing. The denominator reduces to $\gamma^1_h + \gamma^2_h = NL_{h-1} - NL_h$. As $\gamma^1_h = NL_{h-1} - NL_h$, and $\gamma^2_h = 0$, the numerator becomes $NL_{h-1} - NL^{CLP}_{h^*}$ so that we once again have the definition of (4.29):

$$u^{CLP}_{h^*,h} = \frac{NL_{h-1} - NL^{CLP}_{h^*}}{NL_{h-1} - NL_h}\left(t_h - t_{h-1}\right) \qquad \forall h, h^* > 0 \tag{4.34}$$

The development of the definition in (4.33) only considered the cases where either $NL_h \leq NL^{CLP}_{h^*} \leq NL_{h-1}$ or $NL_{h-1} \leq NL^{CLP}_{h^*} \leq NL_h$, however we must also consider those cases where the net load level of interest is entirely above or below the range we are considering. If we consider the case of decreasing net load, then when $NL^{CLP}_{h^*} < NL_h$, the definition (4.34) yields $u^{CLP}_{h^*,h} > t_h - t_{h-1}$, so that

utilisation within the segment exceeds the total duration of the segment. Similarly, when $NL^{CLP}_{h*} < NL_{h-1}$, the expression for utilisation across a segment evaluates as negative. In the former case, we wish to record a correspondence with full utilisation to reflect $NL^{CLP}_{h*} \le NL_h \le NL_{h-1}$ so that the reference net load $NL^{CLP}_{h*}$ is exceeded for the entire duration of that segment. In the second case, we wish to record a zero utilisation, as $NL^{CLP}_{h*} \ge NL_h \ge NL_{h-1}$ so that the reference net load, $L^{CLP}_{h*}$, is not attained within the segment. To bound the evaluation of $u^{CLP}_{h*,h}$ in this fashion we adapt (4.33) into the following complementarity conditions, $\forall h, h* \in H, h > 0$:

$$u^{CLP}_{h*,h} - \frac{\left(\left(1 - \frac{\gamma^1_h}{NL_{h-1} - NL_h}\right) NL_h + \left(1 - \frac{\gamma^2_h}{NL_h - NL_{h-1}}\right) NL_{h-1} - NL^{CLP}_{h*}\right)}{\gamma^1_h + \gamma^2_h}\left(t_h - t_{h-1}\right) + \gamma^3_{h*,h} \ge 0 \;\perp\; u^{CLP}_{h*,h} \ge 0$$

$$\forall h*, h > 0 \qquad (4.35)$$

$$t_h - t_{h-1} - u^{CLP}_{h*,h} \ge 0 \quad \perp \quad \gamma^3_{h*,h} \ge 0 \qquad \forall h*, h > 0 \qquad (4.36)$$

Where the raw utilisation level exceeds the duration of the segment h, (4.35) requires that $\gamma^3_{h*,h} > 0$, which by (4.36) implies $u^{CLP}_{h*,h} = t_h - t_{h-1}$. Where the raw utilisation level is negative, then from (4.35), $u^{CLP}_{h*,h} = 0$. Finally, where $NL^{CLP}_{h*}$ exceeds load only partially within a segment, we have $0 \le u^{CLP}_{h*,h} < t_h - t_{h-1}$ and $\gamma^3_{h*,h} = 0$. In this case, $u^{CLP}_{h*,h}$ must be equal to the calculated or interpolated value or we have a contradiction with $u^{CLP}_{h*,h} > 0$ in which the left side of (4.35) is either negative, which is infeasible, or strictly positive, in which case the complementarity condition is not satisfied.

Having calculated the utilisation contained within each individual segment, then as $t_h - t_{h-1}$ amount to fractions of the day, we must simply sum the utilisation level associated with each reference net load level $NL^{CLP}_{h*}$

$$u^{CLP}_{h*} = \sum_{h=1}^{H} u^{CLP}_{h*,h} \qquad \forall h* \qquad (4.37)$$

The set of points $\{u^{CLP}_{h*}, NL^{CLP}_{h*}\}$ define a net LDC consistent with the net CLP, and with application of the scaling factor defined by S(u), they define a point on the estimated endogenous net LDC.

$$NL_{h*} = NL^{CLP}_{h*} S\left(u^{CLP}_{h*}\right) \qquad \forall h* \qquad (4.38)$$

This endogenous net LDC takes the place of the exogenous load and utilisation levels used in the standard formulation. It remains the case that optimal trade-offs, and their corresponding net load levels need to be determined, and then subjected to the same ordering as detailed in Section 2.6. As before, the ranking procedure integrates and jointly ranks each set of utilisation level, thereby ensuring that the constraints of the model are sensible, while also providing a way to trace between the sorted

ranking and the unsorted ranking of each set of utilisation levels. In this case, this ability is critical, as we must link prices and outcomes with the original generation quantities.

The ordering conditions are:

$$-\mathrm{r}.u^{CLP}_{h^*} + \phi^0_r + \phi^{CLP}_{h^*} \geq 0 \quad \perp \quad x^{CLP}_{h^*,r} \geq 0 \qquad \forall h^*, r \qquad (4.39)$$

$$-\mathrm{r}.u^e_n + \phi^0_r + \phi^e_n \geq 0 \quad \perp \quad x^e_{n,r} \geq 0 \qquad \forall n, r \qquad (4.40)$$

$$1 - \sum_{h^*} x^{CLP}_{h^*,r} - \sum_n x^e_{n,r} \geq 0 \quad \perp \quad \phi^0_r \geq 0 \qquad \forall r \qquad (4.41)$$

$$1 - \sum_r x^{CLP}_{h^*,r} \geq 0 \quad \perp \quad \phi^{CLP}_{h^*} \geq 0 \qquad \forall h^* \qquad (4.42)$$

$$1 - \sum_r x^e_{n,r} \geq 0 \quad \perp \quad \phi^e_n \geq 0 \qquad \forall n \qquad (4.43)$$

The equivalent translation functions are:

$$u_r - \sum_{h^*} x^{CLP}_{h^*,r} u^{CLP}_{h^*} + \sum_n x^e_{n,r} u^e_n = 0 \qquad \forall r \qquad (4.44)$$

$$NL_r - \sum_{h^*} x^{CLP}_{h^*,r} NL_{h^*} + \sum_n x^e_{n,r} NL^e_n = 0 \qquad \forall r \qquad (4.45)$$

$$NL^e_n - NL_0 + \sum_{h^*} \frac{NL_{h^*-1} - NL_{h^*}}{u^{CLP}_{h^*} - u^{CLP}_{h^*-1}} u^{part}_{h^*,n} = 0 \qquad \forall n \qquad (4.46)$$

$$\sum_{h^*} u^{part}_{h^*,n} - u^e_{n,t} = 0 \qquad \forall n \qquad (4.47)$$

$$u^{CLP}_{h^*} - u^{CLP}_{h^*-1} - u^{part}_{h^*,n} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < h^* < H, n \qquad (4.48)$$

The piecewise linear structure that describes the LDC has been replaced with an endogenous piecewise linear net LDC comprise of a chronological net load profile and accompanying adjustment. In this case, the influence of the original LDC specification is limited to assisting the definition of the adjustment function, so that the basic shape of underlying load is preserved as accurately as possible, although as described earlier the original LDC may also be endogenous for reasons such as demand response as discussed in Section 4.2.3. In such cases, the scaling factor function will also be endogenous, to reflect the implications of issues such as demand shifting between periods, for example. The endogenous utilisation and net load levels have replaced the exogenous utilisation and load levels and the value of net load at each optimal trade-off between conventional technologies is interpolated between these endogenous load levels, as it once was between exogenous load levels.

### 4.4.4 Market Clearing

We have decomposed load into a chronological pattern and a scale factor to measure idiosyncrasies in the gross load pattern that are not captured by the chronological pattern. We have then deducted total intermittent generation, which is modelled chronologically, reformed the CLP equivalent LDC, and

applied the estimated scale factor to form a new net load duration curve. Nevertheless, other than modelling the chronological generation pattern and incorporating the necessary adjustments to maintain the integrity of the problem structure, the approach is conceptually the same as the standard deduction approach used to form net LDC's.

Substituting net load for load, and noting that we only consider generation of conventional technologies for servicing net load, the market clearing condition as it applies to the ranked net LDC is:

$$\sum_{i \notin INT} GEN_{i,r} - NL_r = 0 \quad \perp \quad \lambda_r \; free \qquad \forall r,t \qquad (4.49)$$

Depending on the correlations involved, if there is sufficient capacity of intermittent technologies it is possible that $NL_r < 0$. This implies energy spillage and zero load for conventional technologies to service. This situation is depicted in Figure 30.

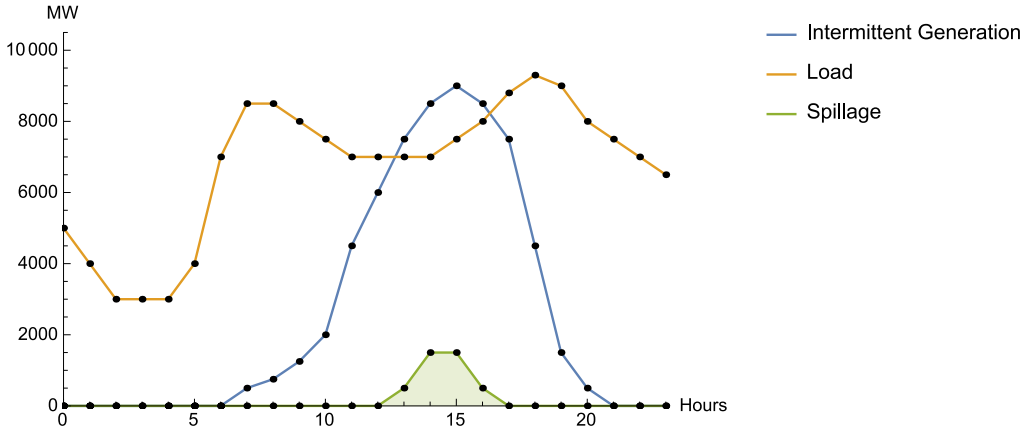



**Figure 30: Energy Spillage**

To accommodate this possibility, we introduce a notional spillage technology with zero fixed cost. Aside from being a wasted resource, we assume that spillage is costless so that the marginal cost of the spillage technology is also zero, thereby providing a floor on the energy price, $\lambda_r$. The market clearing conditions are:

$$\sum_{i \notin INT} GEN_{i,r} - NL_r - SPL_r = 0 \quad \perp \quad \lambda_r \; free \qquad \forall r \qquad (4.50)$$

$$\lambda_r \geq 0 \quad \perp \; SPL_r \geq 0 \qquad \forall r \qquad (4.51)$$

$$-\lambda_r + \mathrm{MC}_i + \varphi^+_{i,r}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r} \geq 0 \qquad \forall i \notin INT, t \qquad (4.52)$$

$$CAP_i - GEN_{i,r} \geq 0 \quad \perp \quad \varphi^+_{i,r} \geq 0 \qquad \forall i > 0, i \notin INT, r \qquad (4.53)$$

When $NL_r < 0$, then since $\sum_{i \notin INT} GEN_{i,r} \geq 0$, (4.50) implies $SPL_r > 0$, signifying spillage is occurring. When $SPL_r > 0$ then from (4.51) we have $\lambda_r = 0$. Conversely, when $NL_r > 0$ we have $\sum_{i \notin INT} GEN_{i,r} > 0$

implying $\lambda_r > 0$ when standard generation technologies have positive marginal costs. From (4.51) we have $SPL_r = 0$, as we should when no spillage is occurring. When $NL_r = 0$, then from (4.50), $\sum_{i \notin INT} GEN_{i,r} = SPL_r = 0$ implying $\lambda_r = 0$.

The market clearing prices generated by the above complementarity conditions are not necessarily chronologically adjacent. For example, two market clearances based on adjacent load levels in the LDC could refer to a morning and an evening output on the chronological load profile. We can determine the pricing pattern in chronological terms by reversing the ranking transformation:

$$\lambda_{h*} = \sum_r x^{CLP}_{h*,r} \lambda_r \qquad \forall h* \qquad (4.54)$$

Similarly, we could estimate the generation of intermittent technology i at a particular point in the chronological load profile as:

$$GEN_{i,h*} = \sum_r x^{CLP}_{h*,r} GEN_{i,r} \qquad \forall i,h* \qquad (4.55)$$

### 4.4.5 Investment

The assessment of any particular intermittent development option can be undertaken by assuming that its decremental impact on the net LDC is sold at the price determined by the corresponding equilibrium PDC. As more intermittent capacity is added to the (hypothetical) plant mix, the shape of the residual LDC will slowly shift, as residual load becomes more sensitive to whether the wind blows, or not. However, in the absence of energy spillage, the shifting of the LDC should not change the equilibrium PDC, which is still based on optimal trade-offs between standard generation technologies. Accordingly, the equilibrium profitability of incremental intermittent investment is not altered. When spillage is possible, investors must consider there is effectively a new technology in the plant mix, which has zero marginal cost, and therefore disrupts the equilibrium PDC. But even when spillage is not an issue, in the presence of existing capacity, investors in intermittent generation technologies must contrast this long run equilibrium perspective with short run analyses, which suggest that increasing intermittent penetration will produce lower returns because prices will be depressed when intermittent generation is high.

We can define equilibrium intermittent generator income using prices defined by the PDC which in turn is defined by conventional technologies servicing the net LDC. Conventional technologies have their profitability determined at each utilisation level and interpolated across operating ranges expressed in terms of those utilisation levels. Assuming marginal costs are zero, in the case of intermittent generation the relevant measure of profitability is the GWAP. To calculate GWAP we need to link the generation of individual technologies with pricing and operating ranges from the PDC. For a single unit of capacity the relevant measure of generation at h* is $ICF_{i,h*}$. Using the intermittent capacity factor and adapting the ranking equation (4.44), we can calculate the generation of technology i at utilisation level $u_r$ using ranking variables:

$$ICF_{i,r} = \sum_{h^*} x^{CLP}_{h^*,r} ICF_{i,h^*} \qquad \forall i,r \qquad (4.56)$$

Accordingly, for each of the r utilisation ranges defined by $\{u_r, u_{r+1}\}$, we have the following expression for estimating the profitability of intermittent generator, which is based on the average capacity factor of the intermittent technology in each utilisation range:

$$\chi_i = \sum_{r<R} \lambda_{r+1} \left( \frac{ICF_{i,r+1} + ICF_{i,r}}{2} \right) \qquad \forall i \in INT \qquad (4.57)$$

Alternatively, we can define the income for each period using the chronological description of the problem, by translating prices from the PDC back to this timescale. The advantage of this approach is that the chronological ordering of generation and prices provides stronger support for the assumption that system performance can be interpolated linearly between adjacent performance measurements. That assumption is more reasonable when the adjacency between measurement points is based on time of day rather than aggregate system conditions, which could result in similar load and price levels at opposite ends of the day which leaves no basis for interpolating outcomes. From (2.86) we can calculate the price corresponding to the boundary of the interval $\{t_{h^*}, t_{h^*+1}\}$:

$$\lambda_{h^*} = \sum_r x^{CLP}_{h^*,r} \lambda_r \qquad \forall h^* \qquad (4.58)$$

$$\lambda_{h^*+1} = \sum_r x^{CLP}_{h^*+1,r} \lambda_r \qquad \forall h^* \qquad (4.59)$$

If we assume that both generation and prices adjust linearly across the relevant time period, we can define a price and generation function that describe their value over the interval $\{t_{h^*}, t_{h^*+1}\}$:

$$\lambda_{h^*}(t) = \lambda_{h^*} + \left( \frac{\lambda_{h^*+1} - \lambda_{h^*}}{\mathrm{t}_{\mathrm{h}^*+1} - \mathrm{t}_{\mathrm{h}^*}} \right) \left( t - \mathrm{t}_{\mathrm{h}^*} \right) \qquad \forall h^* \qquad (4.60)$$

$$ICF_{i,h^*}(t) = \mathrm{ICF}_{\mathrm{i,h}^*} + \left( \frac{\mathrm{ICF}_{\mathrm{i,h}^*+1} - \mathrm{ICF}_{\mathrm{i,h}^*}}{\mathrm{t}_{\mathrm{h}^*+1} - \mathrm{t}_{\mathrm{h}^*}} \right) \left( t - \mathrm{t}_{\mathrm{h}^*} \right) \qquad \forall i \in INT, h^* \qquad (4.61)$$

Given both the generation and price functions are piecewise linear by assumption, the revenue attributable to a single unit of intermittent generation capacity of technology i increases quadratically over the interval $\{t_{h^*}, t_{h^*+1}\}$:

$$REV_{i,h^*}(t) = \left( \lambda_{h^*} + \left( \frac{\lambda_{h^*+1} - \lambda_{h^*}}{\mathrm{t}_{\mathrm{h}^*+1} - \mathrm{t}_{\mathrm{h}}} \right) \left( t - \mathrm{t}_{\mathrm{h}^*} \right) \right) \left( \mathrm{ICF}_{\mathrm{i,h}^*} + \left( \frac{\mathrm{ICF}_{\mathrm{i,h}^*+1} - \mathrm{ICF}_{\mathrm{i,h}^*}}{\mathrm{t}_{\mathrm{h}^*+1} - \mathrm{t}_{\mathrm{h}}} \right) \left( t - \mathrm{t}_{\mathrm{h}^*} \right) \right) \quad \forall i \in INT, h^* \qquad (4.62)$$

The average revenue attributable is therefore:

$$AvgREV_{i,h^*} = \frac{1}{t_{h^*+1} - t_h} \int_{t_{h^*}}^{t_{h^*+1}} REV_{i,h^*}(t)\, dt \qquad \forall i \in INT, h^* \qquad (4.63)$$

The equilibrium investment condition for intermittent generation technologies is therefore:

$$\mathrm{FC_i} - \sum_{h^*} \int_{t_h^*}^{t_{h^*+1}} REV_{h^*}(t)\,dt \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \in INT \qquad (4.64)$$

As with other technologies, investment in a particular intermittent technology cannibalises its own earnings. As the penetration of intermittent technologies increases, the equilibrium PDC decays, with returns to all technologies, including intermittent technologies, being reduced. If the output of intermittent generation technologies is heavily correlated with periods of high net load then this effect is exacerbated. The returns available to the investor critically depend on not only on the output of the generation unit, but also the timing of that output, which cannot be controlled. Preference is therefore given to technologies and locations whose output is strongly skewed towards generating when prices are high, so that the generation weighted price (GWAP) is higher than the time weighted price (TWAP), which would be earned by a generator whose output has zero correlation with system prices. From the perspective of an investor in a particular intermittent technology, the worst case is that the generation is negatively correlated with system prices so that the GWAP is lower than the TWAP.

Although we have not progressed the matter in this work, the possibility of including and analysing the value of storage in partnership with intermittent technologies is an obvious next step. In the limit, where all of the load variation was intra-day, storage devices need only support a fraction of daily output to enable a perfect re-allocation of energy to the times of highest prices.

## *4.5 Summary and Conclusions*

Satisfying load is the fundamental constraint in an electricity market. In this chapter we demonstrated how the framework could adapt to endogenous load. In doing so, we considered load response, reliability and intermittent generation.

In Section 4.2.2., we address short term demand response addresses the practical issue of monitoring, and reacting to electricity market prices in the dispatch. We conjecture that this capability is not widespread and generally it requires investment to gain that capability. Accordingly, the response can be treated as any other technology, in that it is limited in capacity once constructed, and may also be limited in energy, if the response is unable to be sustained, for example. In Section 4.2.3, we progress to long term load response of the kind characterised by load shifting and substitution, either through other fuels or through investment in efficient appliances, for example. By separating the two we confine the role of long term demand response to setting the underlying LDC, which is modelled as being completely inelastic in the short term as short term responses are considered technologies. Under this approach, the underlying LDC is endogenous and dependent on average and relative prices. In market clearing, this demand response cannot be marginal, as market clearing operates on the endogenous LDC, not the original LDC. We achieve this by adding complementarity conditions to define a further technology, whose capacity is equal to the demand response, and whose marginal cost is set to ensure it will always be utilised.

In Section 4.3, we adopted the conventional augmented load approach to addressing reliability. Given a capacity mix at any particular generation level we can calculate the expected

impact of less than perfect reliability, in terms of the additional capacity required, and add that to the LDC. From that point we assume capacity is perfectly reliable. In that respect, we calculate material outages only, noting that the marginal generating technology may itself have sufficient idle capacity to cover losses. Whether the reliability of each technology results in increased capacity relative to other technologies depends on the relative reliability of those technologies. However, low capital cost technologies will fare relatively better than high cost technologies, although we note there is probably a more fundamental relationship between cost and reliability that makes skews that comparison. In the limiting case, the notional shortage technology is completely reliable, and therefore shortages will be more prevalent when technologies are unreliable. In terms of our framework, the redefinition of the LDC naturally changes the numerical solutions obtained, but the definition of optimal trade-offs requires no adaptation. The complementarity constraint governing investment is adjusted for reliability so that depending on the perspective, returns are diminished to account for periods where a unit is offline, or the cost of perfectly reliable capacity is higher.

Finally, we investigate the inclusion of intermittent generation in the framework. Our particular motivation was capturing the chronological correlations between intermittent generation and load directly. Our approach is conceptually straightforward. We begin with three sets of data: the LDC, a chronological load pattern, and a chronological generation pattern for each intermittent technology. We define the relationship between the chronological load and the LDC and develop a scaling factor to reflect the additional variability present in the LDC. We are then able to determine net load by deducting intermittent generation sources form the chronological load pattern, taking account of the potential for energy spillage as we go. In doing so, we create the need to define a representation of net chronological load in the form of a LDC. The resulting net LDC is then rescaled and used as the basis of market clearance and generation by conventional technologies. In equilibrium, the pricing that results is also paid to intermittent generators and this disciplines investment.

The complementarity constraints that perform this task are complex. While the scale factor will be decreasing, it may not be precisely monotonically decreasing. To define a monotonically decreasing scale factor we propose a constrained cubic curve fitting based on a least squares objective, the parameters of which are restricted to ensure the curve is monotonically decreasing. The KKT conditions from this optimisation form part of the complementarity formulation. Next, we must consider the basic construction of an LDC from a chronological load pattern. To do so requires determination of the utilisation level of load from that pattern. Given a load level, we introduce complementarity conditions to define across a segment the extent to which load is above or under the load level of interest. This procedure is applied to net load, which must account for spillage of energy in the case when intermittent generation exceeds underlying load. To achieve this we introduce an additional costless technology. Finally, it is necessary to determine the earnings of intermittent technologies, and for this we determine the market clearing price at a series of points along the chronological load profile, and interpolate the generation between points to get an average capacity factor across the intervening segment.

The contribution of our approach to reliability, while slightly nuanced on account of addressing only material outages, limited to demonstration of the inclusiveness of our framework.

With respect to demand response, while we have seen models augmenting load, and other models considering demand response as a technology, we are unaware of any implementation of a specific approach such as ours, that separates long term and short term response in a conventional optimisation formulation of electricity markets. As with the basic model, these models can certainly be conceptualised as optimisation formulations, but the determination of optimal trade-offs and consistent solutions means these problems cannot be solved with a conventional optimisation formulation. We are also unaware of any investment model that incorporates both and chronological load patterns with an LDC representation. We do this and go beyond by considering the above relationship in the context of net load, constructing a net LDC dynamically, that is then incorporated into our wider framework to ensure the solutions are consistent. The complementarity conditions that define the relationships required to implement this are entirely original.

# 5 RISK & UNCERTAINTY

## *5.1 Introduction*

In Chapter 4 we considered several forms of variability, such as intermittent generation and reliability. In these cases the variability in the system is material, particularly when supported by correlations, but ultimately the influence of variability for risk neutral investors is limited to its influence on expected returns. In this chapter we introduce risk and uncertainty to the framework, after which investors are no longer interested in expected returns alone, but also the distribution of returns. Our goal is to explore how risk and uncertainty can be implemented in the framework we develop.

We begin by reviewing the definition of risk, uncertainty and other related concepts. In particular, we reiterate a long known, but often overlooked, distinction between risk and uncertainty. The former is defined as variability with known probabilities, while the latter is variability with unknown probabilities. After detailing some of the evolution of risk measures and the advantages and disadvantages of each we elect to implement CVaR, it being a coherent and mathematically tractable risk measure. A number of paradigms have been used in risk management. Perhaps the most famous is the portfolio construction approach, and the CAPM model that followed it. In more recent times, the concept of portfolio replication has become popular, particularly in financial markets where the replication and therefore hedging of a portfolio is efficiently achieved because the large number of financial assets available. Finally, in this respect, we consider a new paradigm in risk modelling: the stochastic endogenous equilibrium, in which the price of risk is also a variable.

The optimisation of CVaR is well documented but less so when in convex combination with expected returns. While a definition based on the exclusion of some proportion of the best scenarios can simulate the convex combination, we choose a definition of CVaR that is oriented to the risky end of the distribution of outcomes that we combine with expected returns. Our direct formulation of CVaR is consistent with other formulations, but our direct implementation of the dual of the traditional formulation in a complementarity framework is efficient and, to our knowledge, original. As others have done (Ehrenmann & Smeers, 2011), we show that our risk measure generates a set of risk adjusted weightings that provide an alternative to the objective scenario weightings in the model. The inclusion of risk necessarily affects the marginal benefit of investment and we explore those impacts in general and with some examples. The approach taken is extended to include nuanced CVaR preferences and CVaR constraints, which can be defined at multiple significance levels, or with respect to subsets of the scenario tree.

While the above addressed a portfolio approach to capacity investment, contracts are also available. We consider the formulation of contracts and the integration of contracts with the CVaR risk measure. We do so in the context of a simple forward contract, designed to protect the contract holder from price risk. Rather than have contract prices fixed, we consider the structure of contract supply and demand directly, and then implement market clearing conditions. In doing so we are able to

explore the results of the model in the context of a stochastic endogenous equilibria where risk is priced in the contract market endogenously (Ralph & Smeers, 2011).

Our formulation of uncertainty is exploratory. Uncertainty is used to describe those forms of stochastic variability for which no probability distribution can be formed. We posit an approach to modelling uncertainty based on the concept of utilisation and consider the effect of uncertainty on the framework developed so far.

## *5.2 Conceptual Framework*

### 5.2.1 Variability

Variability does not, of itself, indicate the presence of risk or even stochasticity or randomness, as it can be deterministic, stochastic, or both. There are many examples, such as in seasonal industries, where a great deal of variability is present, most of which is in large part predictable and does not represent risk. The most important characteristic of variability in our modelling context is the nature of correlation between different forms of variability, its relationship to risk, and how it is best incorporated in an investment model.

In the first instance, variability is important because of correlation. Even when predictable variations occur, it is important to understand the correlation between those variations, or alternatively the conditional proportions associated with each. It may be, for example, that the fuel cost and the level of demand are correlated seasonally, magnifying the effect of each on profitability.

From the perspective of a modeller, the distinction between variability and risk is important, as a risk measure should not penalise any portion of an outcome that is due to predictable variations in the model. For example, the predictable portion of the seasonal load variation does not reflect risk, although the distribution around that predictable fluctuation might do. So, for example, typical high prices in a summer peaking market should not have risk implications, whereas high seasonally adjusted prices might well do.

Finally, in a modelling environment with all of the standard foibles, the relevant practical distinction may be whether variability or randomness is modelled deterministically or stochastically, rather than whether the variability is actually predictable or stochastic. In this framework, the LDC represents the stochastic variability of load in the form of the LDC, which is a proportional form. The implication of the LDC is that for the period to which it relates we assume that all load levels represented will be experienced in the proportion shown. This is subtly different from considering the continuous sampling from a load distribution, that might well coincide with the LDC. We then considered other forms of variability that can be described through adjustment or amendment to the LDC. These included reliability, and through a somewhat more detailed process, the impact of intermittent generation. In this chapter, we explicitly consider scenarios with associated probabilities, the outcomes of which are subject to inclusion in risk measures designed to represent investor preferences.

*Variability in Electricity Markets*

Electricity markets contain many sources of variability. A survey on risk management in electricity markets is available from Liu et al (2006) and includes many sources of variability as motivation. Weber (2011) also offers a comprehensive treatment. Other than the issues discussed explicitly (plant reliability, intermittent generation, and hydrology or more generally fuel inflows, the following non-exhaustive list notes some of the more significant examples of variability that has implications for risk and uncertainty:

- Demand Growth. While a source of variability and of relevance to investment timing (Bean, Higle, & Smith, 1992), this is not something we have focussed on. In an equilibrium model, the effects of reduced demand growth, for example, should be as fleeting as the equilibration adjustment. We note that the efficacy of the adjustment process is not beyond doubt, and that the most prevalent issue facing investors is typically net demand growth, after accounting for renewables.
- Fuel cost and fuel supply. This is a significant issue, particularly when connected when worldwide fuel markets are interconnected and fuel prices have the potential to vary significantly on the basis of global adjustments in fuel markets. In areas where energy markets and fuel supplies are prone to disruption there are a number of fuel supply concerns (Boucher & Smeers, 2012), and there can be significant merit order risk, with potential for cycling (Bell, 2010).
- Taxation and Climate Policy. Taxation is an important consideration for all investors as it reduces returns. In recent times, taxation has also become a mechanism for adjusting relative fuel prices to achieve environmental goals. Other forms of climate policy based interventions, such as the introduction of carbon trading, are also possible and the uncertainty surrounding their introduction or evolution is an important factor in the investment decision (Blyth, 2007), (Kettunen, 2008).
- Market Intervention. Electricity markets are both young and technologically dynamic, and this has created a tendency for regulators and governments to intervene from time to time to right perceived faults in market designs.
- Competitor Strategy. As strategy, such as spot market gaming, alters returns, it also directly modifies risk or uncertainty adjusted returns. Variability in strategy also leads to strategic risk and uncertainty. In the case of risk, the strategy may be known up to a distribution surrounding some parameter values such as costs, whereas in the case of uncertainty, the competitor strategy is unknown.

The relative importance of each aspect of variability is market dependent and modellers should reflect the typical concerns of an individual market by prioritising those items requiring the most focus.

### 5.2.2 Risk

Several definitions of risk exist. There is a colloquial definition of risk, which defines risk as "exposure to detrimental outcomes". The colloquial definition is not without informal support in the academic literature (Fishburn, 1984), although this support is often not as much a specific endorsement

of the colloquial definition as it is the use of a convenient term used to describe a commonly understood principle.

For the purpose of analysing the concept of risk, the distinction is not important, but when the analysis turns to the specific and distinct concept of uncertainty, that definition is unsatisfactory. The colloquial definition of risk actually coincides more precisely with terms such as "Downside Risk", which are also in common use, and in which the term "Downside" would be redundant if it were used in conjunction with the colloquial definition. In all but the case of symmetric payoff distributions, this definition also conflicts with long-standing technical measures of risk such as variance, which also reflect the possibility of exposure to superlative outcomes. However, from the perspective of this research, the most unsatisfactory feature of the colloquial definition of risk is that it does not lend itself to a useful distinction between risk and uncertainty.

We adopt a precise and analytically useful definition of risk, originally offered by Frank H. Knight (1921), in which risk was defined as "randomness with knowable probabilities" and uncertainty was clearly distinguished as "randomness with unknowable probabilities". Much of the literature comports with this definition of risk (Farrar, 1964), (Alessandri, Ford, Lander, Leggio, & Taylor, 2004), insofar as agreeing that risk measures are defined over distributions. However, few authors distinguish between risk and uncertainty and many, for example (Genc, Reynolds, & Sen, 2007), (Shanbhag, Infanger, & Glynn, 2011), (Conejo, Carrión, & Morales, 2010), use the terms interchangeably.

### 5.2.3 Risk Aversion

In an equilibrium model, we should only be interested in risk if investors are themselves interested. Investors may be either risk seeking, risk neutral, risk averse, or a combination of these, adjustment their attitude on the basis of wealth or the size of the gamble they face, for example. Risk aversion describes what is typically assumed to be the first order reaction to risk; that being we do not like it. The degree to which investors are risk averse describes how far they would go to avoid risk or, alternatively, the rate at which they are prepared to sacrifice expected outcomes in order to increase certainty. Risk neutral agents are only in the expected outcomes, while risk seeking agents actively seek riskier portfolios at the expense of expected outcomes.

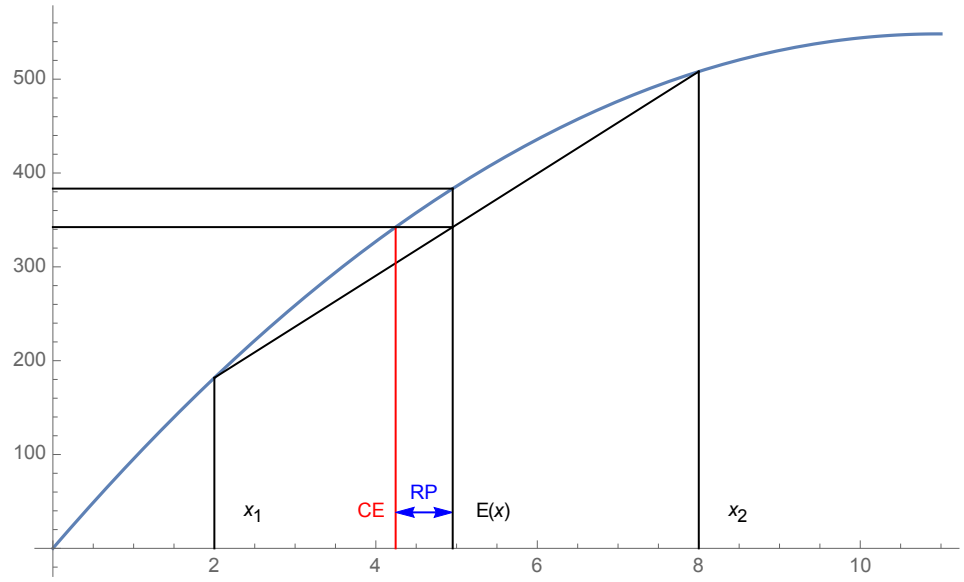



**Figure 31: Risk Aversion and Risk Premiums**

*Source: Wolfram Demonstrations Projects, John Horton*

To the extent that an investor is risk averse, they will be happy to trade off a portion of the expected value of their profits, asset value or other objective, as consideration for facing less risk. The value at which the investor is indifferent to the gamble on offer is known as a certainty equivalent. The relationship between expected returns, the certainty equivalent, and the risk premium is shown in Figure 31, in which E(x) represents the expected value of outcomes, CE is the certainty equivalent, and RP is the risk premium.

Risk aversion can be decomposed further, into increasing, constant, or decreasing, absolute or relative risk aversion. We do not address the implications of risk aversion being a function of firm size, but we note that when risk aversion is a decreasing function of firm size, an economy of scale exists, which has implications for the industry structure.

### 5.2.4 Uncertainty

From Knight (1921), the definition of uncertainty is "randomness with unknowable probabilities". Risk and uncertainty are distinguished by whether the distribution of the variability is known, as is the case with risk, or not, as is the case with uncertainty. A strict interpretation suggests there is only uncertainty, as no distribution of interest to investors is known with certainty. However, in many cases the distribution of certain forms of variability are well, if not perfectly, understood and it would not be appropriate to overlook this information on the basis it was not absolutely definitive (Farrar, 1964). In such cases the variability can be decomposed into the systematic which can be addressed as risk, and the esoteric, which remains uncertainty.

We also note that investors are not simply concerned with the accuracy of input distributions such as for hydrological conditions, demand forecasts or solar energy availability. Investment

decisions are fundamentally based on output distributions such as the equilibrium PDC, and this brings issues such as the validity of the model specification or structure, or even whether the equilibrium PDC is the appropriate basis for decision making into consideration. So, the difference between risk and uncertainty is more than a theoretical demarcation along a continuum of descriptions of variability bounded by pure risk and pure uncertainty at its extreme. The reason for the distinction is to separate risk and uncertainty, as investors do in their decision processes.

By its very nature, risk requires the knowledge of distributions, and that in turn requires either theoretical justification of distributions based on probabilities, or the accumulation of sufficient data so that the sample distribution may be relied upon as a population distribution. In order to accumulate sufficient knowledge of a distribution, the sample size must generally be large, and therefore observations must occur with high frequency. Where theoretical distributions do not exist or observations are too infrequent to establish an empirical distribution, the variability must be classified as uncertainty. In the context of electricity markets this suggests that variability on a daily basis in the spot market can be treated as risk at worst, but questions of government policy, or the long term rate of climate change, are best described as uncertain, given the underlying model driving each is not clear, and the availability of data is limited.

This conclusion is also supported by the structure of contract markets in many electricity markets. Contracts are generally based on a sound understanding of the risks involved. Contract markets are not likely to develop in response to uncertainty. A lack of knowledge of relative probabilities makes it difficult for a counter-party to assess uncertainty and value contracts. In the limit, it is difficult to satisfactorily assess the likelihood or impact of an event that has never occurred, let alone one that has not even been considered. Where long term contracts that are subject to significant uncertainty are offered, these might be justified by strategic interests and are likely to be struck bilaterally by parties wishing to neutralise the impact of uncertainty. Consequently, we are not aware of these types of contracts being exchange traded in a form that a more diversified independent party, such as a financial institution, could trade in.

## *5.3 Risk & Uncertainty Management*

### 5.3.1 Introduction

Risk management refers to the method used to control risk. These methods extend to identifying risk, assessing risk, measurement of risk, and mitigating risk. This can occur at many levels of the firm, and can be rule based, or more flexible. Risk management occurs at a variety of levels within a firm, and address different timeframes. A typical view of the time and decision structure applicable in electricity markets is provided, along with an assessment of the relative balance between risk and uncertainty is provided:

| Horizon | Timeframe | Risk/Uncertainty Balance |
|---|---|---|
| Investment | 20-30 years | Mostly uncertainty |

| | | |
|---|---|---|
| Medium Term | 2-3 years | Balanced |
| Seasonal | Season Length | More risk |
| Trading | 1-2 months | Mostly risk |
| Spot Market | 48 hours | Completely risk |
| Regulation | 1-2 hours | Completely risk |

**Table 4: Decision Timeframes in Electricity Sector**

The length and nature of these horizons depend on the specifics of the system under consideration. Our focus is on the investment horizon corresponding directly to the lifespan of the investment, and the medium term which corresponding to timeframes ranging from seasonal periods through to several years depending on the duration of underlying variability, as decisions on these timeframes materially impact the distribution of annual returns.

### 5.3.2 Flexibility

Flexibility underlies risk management. The concept of flexibility, which may be thought of as a capability to react and respond to unforeseen events in a variety of ways, determines the need, or extent to which a firm or investor need go to manage risk (Ku, 1995). Flexibility can be a function of the firms portfolio as well as the decision making process itself, which, depending on the relative importance attached to flexibility, can prioritise strategies that promote more or less flexibility, ultimately at the expense of other objectives.

Flexibility can also be a source of competitive advantage, and is subject to valuation in a portfolio context (Doege, Schiltknecht, & Lüthi, 2006). The implication of flexibility and the time it will take a firm to respond to a particular eventuality, is that, ceteris paribus, investors will value options with greater flexibility higher than those with less. This suggests that firms may be prepared to contemplate accepting larger levels of short term risk, provided the available recourse actions and means of mitigation enable the risk taken to be easily hedged at a future date. An example of flexibility in this work exists on the demand side, where the flexibility to respond to electricity prices in close to real time may or may be cost prohibitive, but is nevertheless valued and impacts on the desirability of contractual protection.

From the supply side, a most notable example of the value of flexibility in electricity markets is the operational risk management conducted by hydroelectric producers with storage. Risk aversion further complicates hydro scheduling, as does the interaction with contracts (Barroso, Granville, & Trinkenreich, 2003), (Wallace, 2009). As the hydrological conditions for a particular season unfold, hydro producers continuously hedge their position based on a re-assessment of future prospects.

### 5.3.3 Measuring Risk

*Variance*

The most commonly statistic used to summarise the spread or risk of a distribution is the variance of the distribution. Unfortunately, where the distribution of outcomes is not symmetric, the applicability of variance as a risk measure is questionable (Francis & Archer, 1979). When the return distribution is

asymmetric the minimisation of variance as the objective function can lead to selection of stochastically dominated investments simply because they have lower variance (Blavatskyy, 2010). Appendix 7.6 describes stochastic dominance.

This is precisely the case in electricity markets, where return distributions are not symmetric. For example, if we consider investment in a peaking technology in a hydro dominated system, the return distribution will be significantly skewed, reflecting a predominance of years with little or no return in which hydro inflows were sufficient enough to enable the effective management of storage to avoid the need for peaking plant to be used. In the odd year, when hydro inflows fall short, the energy shortage will necessitate generation by the peaking technology. These events will be relatively rare, although the technology will be highly profitable, and a significant proportion of lifetime profits will be earned, at these times.

*VaR*

In response to the shortcomings of variance and in concert with increased computational ability, additional and more complex risk measures have been developed. Many firms and investors are not concerned with variability on the upside of the return distribution as they are with variability on the downside. Therefore, in terms of the distribution of returns they face, their interest lies in the tail containing the worst outcomes.

Value at Risk (VaR) provides an alternative risk measure (Charnes, Cooper, & Symonds, 1958), (Prékopa, 1973). In our context, VaR represents the minimum loss, or best outcome, incurred in the worst $\alpha\%$ of scenario outcomes. The downside of this approach is that it does not consider the nature and size of the losses that lie beyond the threshold. Those scenario outcomes could represent incrementally larger losses, or they could represent losses that would pose an existential risk to the firm. VaR provides no way of distinguishing between these cases, so that given two return distributions with equal losses at the 5% level, where one distribution has catastrophic losses at the 1% level and the other does not, the VaR measure is indifferent.

Although we do not consider such examples, there are occasions when the nature of a particular risk might be non-convex and therefore suitable for a VaR type approach. In the context of investment, outcomes worse than VaR might apply to company failure, for example, beyond which there are no meaningful degrees of failure.

The definition of VaR was helpful in that it focussed attention on a subset of the distribution that aligned with our colloquial understanding of risk. However the failings of VaR as a risk measure are significant. As Artzner (1998) notes, perhaps the most aggregious issue with VaR is that it fails a basic sub-additivity test, so that the risk of a combination of risks may be greater than the sum of the individual risks. As they note, the implication is that an investor in equities, for example, would be incentivised to operate two accounts, carrying out a single investment in each. In doing so they could end up with less risk than the counterparty who may hold the risks together. The implication is that when using VaR as a risk measure, diversification could be relatively discouraged.

*Coherent Risk Measures*

To improve the conceptual basis of risk measures, Artzner (1998) defines the properties of "coherent" risk measures. Here r is a coherent risk measure, and X is a portfolio

- Translation invariance. When a is a deterministic portfolio:

$$r(X+a)=r(X)-a$$

  Adding a riskless asset to your portfolio reduces your risk by the value of that riskless asset.

- Subadditivity. Where $X_1$ and $X_2$ are two portfolios, we have:

$$r(X_1+X_2)\leq r(X_1)+r(X_2)$$

  The risk of two portfolios added together cannot be greater than the risk of each portfolio added. This principle underpins portfolio theory, and is otherwise known as the diversification principle. The degree to which the risk of the combined portfolio is less than the sum of the risk of each separate portfolio is determined by the correlation of the portfolios. VaR fails the sub-additivity test.

- Positive Homogeneity:

$$If\ \alpha \geq 0\ then\ r(\alpha X)=\alpha r(X)$$

  Multiples of a portfolio, multiply the risk of the portfolio, implying the risk of a portfolio is proportional to its size. This also follows from the diversification principle, as it applies to the specific case of perfectly correlated portfolios being added.

- Monotonicity. Where $X_1$ and $X_2$ are two portfolios we have:

$$If\ X_1 \leq X_2\ then\ r(X_1)\geq r(X_2)$$

  Where one portfolio (weakly) dominates another in all scenarios, then that portfolio should have lower risk on account of having higher returns. This is not the case with variance based measures where a portfolio with higher returns might also have higher variance and be considered riskier.

The top three properties define convex risk measures, and in combination with the last we have a definition of the properties of coherent risk measures.

*CVaR & Downside Risk*

Conditional Value at Risk (CVaR) was introduced in Rockafellar & Uryasev (2000) which also details the evolution of risk optimisation from VaR to CVaR. In addition to the theoretical issues noted by Artzner (1998), they also note that VaR is non-convex and difficult to optimise. CVaR has many desirable properties, especially when compared to VaR. Most importantly, CVaR is a coherent risk measure (Pflug, 2000). VaR can be coherent Unlike VaR, CVaR also satisfies second order stochastic dominance (Ogryczak & Ruszczyński, 2002) giving it a degree of credibility that variance based measures, for example, do not have.

Sarykalin (2008) provides an excellent summary of the comparative advantages of CVaR and VaR. They conclude that VaR is computationally difficult to optimise, as it is a non-convex measure. In contrast, CVaR has significantly easier mathematical properties and, by virtue of considering all scenarios in the relevant tail of the distribution, CVaR controls scenarios beyond the level of VaR. Although Pang and Leyffer (2004) provided a complementarity based VaR minimisation formulation, Rockafellar and Uryasev (2000) had already presented a CVaR optimisation that can be formulated as a much simpler linear program.

CVaR represents the expected losses in the worst $\alpha\%$ of scenario outcomes and therefore captures information about the distribution beyond the threshold of VaR.

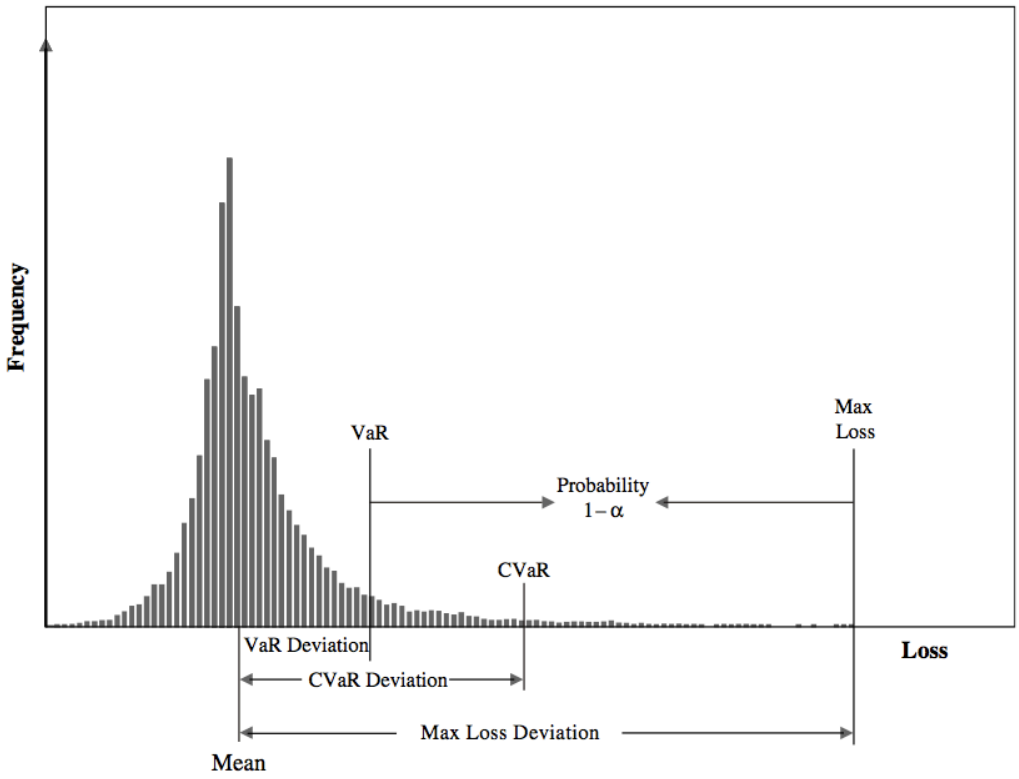



**Figure 32: VaR, CVaR & Downside Risk**

*Source: Sarykalin et al (2008)*

Figure 32 also shows some alternative deviation measures that are based on VaR or CVaR and the deviation from either the mean or the maximum loss. Downside Risk is an example of a deviation measure, as it is the difference between a specified profit level and an actual profit level, as achieved in adverse conditions. Where it is defined with respect to the expected profit, downside risk coincides with CVaR deviation as shown, but downside risk need not be defined relative to the expected profit, as there may be some other profit level, such as the profit level corresponding to zero net cash flow, that could define the point at which adversity might begin to create compounding losses.

*Suitability of CVaR*

The profitability distribution will not likely approximate the normal distribution, or any other symmetric distribution, so the applicability of variance as a risk measure in this case is questionable. As noted, variance is a far less precise measure of risk with asymmetric distributions as it encompasses information from the whole distribution, equally weighting the sensitivities that are associated with each tail of the return distribution, despite one of the tails being more relevant from a risk perspective. In contrast, CVaR can be focussed on the most relevant parts of the distribution, where risk is concentrated, and it is for this reason, and because of its theoretical properties of convexity and consistency, we choose CVaR as a risk measure. The downside of the relative responsiveness of CVaR to changes in the tail of a distribution of outcomes is the magnification of the error that results when that distribution is poorly specified. There are two issues of concern here; the degree to which relevant input distributions and any potential correlations are understood, and the degree to which the formulation translates these inputs into modelled outcomes on which risk measures can be reliably based.

We expect that those items that are identified as risks will, by definition, have well understood distributions and joint distributions, on which analysis can be based. Indeed, where the nature of random variations cannot be represented reliably with a distribution we have a prima facie case against the application of a risk measure at all, and might suggest that an approach based on uncertainty may be more appropriate. For example, a modeller, with limited powers of investigation, is unlikely to ascertain the probability distribution of a competitor's strategic response or government action either now or at some time in the relevant future.

Whereas a modeller may not be in a position to accurately define an input distribution, the accurate solution of the model lies within the modeller's sphere of influence to a much greater extent. As CVaR is assessed on the distribution of returns, it is particularly sensitive to errors in the specification of the PDC. Unfortunately, as was shown in Chapter 1, conventional optimisation formulations are potentially flawed in this area. These models require the contradictory assumption of non-competitive pricing to establish an equilibrium, and the arbitrary restriction of generation functions based on the LDC definition results in unquantified errors in the assessment of prices and price durations.

The proportion of time in which the system is short of capacity, in which prices typically rise to very high levels, is often key to the definition of the tail of the profit distribution in investment problems. Our approach, outlined in Chapter 2, determines the fraction of time the system experiences shortage endogenously. By extension, and by virtue of defining market clearances at endogenously defined points, the PDC is also determined so that, given the input data, not only are expected earnings precisely defined, no additional error is introduced to the calculation of CVaR as individual sub-period PDC's are precisely defined and therefore the tail of the return distribution is also precisely defined.

Targeted increases of the granularity of the model may be a possible solution strategy in a simple single scenario case, but increasing modelling granularity is far more problematic in a multiple scenario case, as to ensure an accurate result the modeller must assess the utilisation levels corresponding to optimal trade-offs in not just a single scenario, but in all scenarios. With respect to

the conventional optimisation formulations, it is unclear a priori how many utilisation levels need be considered to approach an acceptable, or indeed any, accuracy level. It is also difficult to target such utilisation levels at the zone applicable to the shortage calculation, as this varies significantly by scenario, or by sub-period, and to a degree that is a function of the overall equilibrium and therefore is difficult to understand without first solving the problem.

By failing to correctly define the PDC, standard investment models mis-report earnings in each scenario and therefore mis-report the distribution of earnings. In addition, the sensitivity of those earnings estimates is increased by the focus on risk. Juxtaposing the inaccuracy introduced by restricted generation functions with the sensitivity of CVaR to specification of the tail of return distributions, a modeller using a conventional optimisation formulation might well reasonably decide to adopt variance as a risk measure. By being less focussed on the tail of the return distribution, variance minimises the extent to which outcomes could be subverted by an incorrectly specification of that distribution due to a modelling anomaly. In contrast, our approach, which is based on the use of endogenous utilisation levels, improves the definition of the PDC and makes the use of CVaR less susceptible to error.

*CVaR Implementation*

When implementing CVaR, the modeller has some choices, including which portion of the distribution to focus on, and whether CVaR is to be used in conjunction with a measure of expected outcomes. We first consider the use of a lone risk measure as an objective. Ehrenmann & Smeers (2011) details this exact approach, in which CVaR is defined by calculating the expected measure across all but the most favourable scenario's, suggesting a "negative upside risk" approach, rather than a downside risk approach. The inversion of the CVaR intuition is designed to address a wider range of outcomes, in an effort to give positive weighting to these. By orienting the formulation in that fashion, the mean-risk decision is simulated, albeit coarsely. In contrast, maximisation of a similar CVaR measure focussed on actual risk management concerns such as wet/dry years or cashflow management would lead to an outcome with investment supported only to the level at which it is profitable, even the worst scenarios.

Unfortunately, this approach has two disadvantages. Firstly, while the approach implicitly recognises the role of central or median outcomes in the investment decision in a model, it ignores a portion of the distribution altogether, in defiance of economic and financial logic. The implication being that, for example, two otherwise equal bets with different windfall payoffs, are valued identically by investors.

Secondly, ex ante, the risk measure has no intuitive link to actual risk management or the underlying principle of downside risk. While this approach can potentially be tuned to give a numerically equivalent result, it is unclear how one would select which portion of the distribution to focus on, in order to achieve the appropriate relative avoidance of certain critical downside risk demarcation points. The problem with a lack of linkage to ex ante risk management is that this is the basis on which preferences are likely to be defined. The approach taken would require iterative solution and adjustment of the CVaR measure in order to attain consistency with the actual risk management objective of the firm.

As the authors note, further investigation would involve combining risk measures with expected earnings, and that is the approach we have taken. Our application of CVaR aligns naturally with the conceptual basis of risk aversion by calculating the CVaR measure across only unfavourable scenarios. The creation of a convex combination of returns and risk measures enables the specification of risk measures that align with the risk management issues being faced while preserving consideration of the entire return distribution, making the result sensitive to all scenario outcomes. This, more consistent, approach facilitates the consideration of multiple risk measures, which can be designed to specifically address different levels of sensitivity or particular subsets of the scenario tree.

### 5.3.4 Risk Management Paradigms

*Portfolio Theory*

Markowitz (1952) introduced a portfolio approach to investment analysis in which exploration of the possibilities of risk and return combinations were central. Motivated by the observation that investors held diversified stock portfolios that were not explainable by the risk-neutral maximisation of expected returns, this seminal paper shows how consideration of the co-variance of returns between individual assets leads to portfolio diversification. Depending on how implemented, Markowitz's approach minimises return variance subject to a return constraint, or vice versa. Solving this optimisation parametrically by varying the rate of return requirement produces an efficient frontier that describes the Pareto-efficient options available to the firm. Based on their preferences, the investor can then select from this set of non-dominated options the most desirable portfolio. Conceived with equity and financial markets in mind, the concept of variance and co-variance between returns of different types of assets was primarily a statistical concept, reflecting issues such as the counter or pro-cyclical nature of certain industries or firms within industries. The use of variance has often been the subject of criticism as it implies symmetry in the distribution of asset returns that is not necessarily realistic. The assumption of symmetric returns is particularly difficult to support when considering electricity generation technologies, and becomes increasingly untenable in single payment markets as we consider technologies operating higher in the merit order.

*CAPM*

Sharpe (1964) extended the basic approach of Markowitz and developed the Capital Asset Pricing Model, or CAPM as it is commonly known. The CAPM is based around a risk measure known as beta, which measures the contribution of an individual asset to portfolio risk. Marginal risk is measured by the increase in risk to the whole portfolio, and is not a simple function of the riskiness or variance of returns for that particular investment. The effect on the portfolio is of central importance as a significant portion of research into investment decisions treats investment decisions as standalone decisions. Importantly, beta is exogenous for investors in this framework.

Consideration of the entire portfolio is very important in the context of electricity investment. In the electricity industry. Investors face a different issue to equity investors as contracts, existing plants and new investments will have complex interactions under a variety of different scenarios. The interaction of investments within a portfolio of generating assets is dependent on some random variability, but there is also an element of control that can be exerted in terms of operational strategy

that will modify the relationship between asset returns, that is not available to investors in equity markets. While the CAPM model has a number of detractors based on, for example, its reliance on variance, it remains widely used as it is convenient and has the benefits of familiar consensus in financial markets (Jagannathan & McGrattan, 1995)

The implied diversification of assets in the Markowitz/CAPM models is not based on the traditional but naive mantras of "not putting your eggs in one basket", or in "playing it safe". Markowitz/CAPM diversification chooses specific diversifications, which may, if negatively correlated, be based on just two assets, both of which are individually assessed as "risky". Both the position of the generator and the market value of a contract written on spot prices are individually risky propositions, heavily reliant on several factors. They are also negatively correlated, making each an ideal risk management tool for owners of the other. This diversification of risk forms part of the basis of contracting markets, but even when contract markets are ill formed or incomplete, vertical integration provides another form of diversification that reduces portfolio risk.

*Portfolio Replication*

Portfolio replication refers to matching the returns of a particular investment with other financial instruments. The replicated portfolio can be either sold or bought to hedge the value of the original investment by precisely cancelling out the variability of returns that flow from the original asset. Accordingly, portfolio replication is a useful risk management paradigm. Portfolio replication is seldom perfect although the quality of the replication will improve with the number of traded assets. There is also the question of transactions costs, particularly where the number of traded instruments required to hedge the portfolio is large, or where the portfolio is dynamically hedged and requires frequent updating. Nevertheless, if the market is a complete market, portfolio replication provides the investor with a way to end up in a risk free position (Ralph & Smeers, 2011).

*Stochastic Endogenous Equilibrium*

Statistical diversification and portfolio replication are the two classical methods for managing risk (Ralph & Smeers, 2011). In each case, the prices of all instruments are known. In the case of portfolio replication, the PDF describing future asset price movements is assumed to be known as is the variance of future price movements in the case of Markowitz portfolio theory. These data are derived from empirical studies and not from the clearance of an endogenous market. Accordingly, the actions of investors in these markets have no influence on prices, and therefore no influence on each other, and are able to solve their own optimisation problem independently.

In a complete risk market, there is a risk neutral PDF in which the price of every asset is the expectation of its payoff. The basic complementarity formulation that represents the investment and generation problem and that runs throughout the thesis is an example of a risk-neutral Nash game based on risk-neutral probabilities. The actions of each participant impact the other and the equilibrium is found at a point where no participant has an incentive to adjust.

If we assume that agents are risk averse we can incorporate a risk penalty into their objective function, and given a fixed price of contracts, we can also optimise the contractual protection they seek. We do this in Sections 5.4 and 5.6. Again, an equilibrium can be found using the set of adjusted

objective functions however we do not follow that approach other than for the purpose of developing the objective function of the market participants.

Instead we follow the approach of Ralph & Smeers (2011,2015). They add to the model structure above by determining the price of risk endogenously, defining clearance of the contract, or more broadly, the risk market. The *risky design equilibrium problem* adds market clearance of a complete risk market to the set of individual risk-averse objective functions. This represents a risk averse Nash game with risk market explicitly modelled.

Each participant has a risk set, D, which describes the risky scenarios or events that they may experience. Where the agent uses a coherent risk measure we can describe the risk measure as being the worst value of the objective function over that set, where the objective function is usually cast in terms of expected profit or loss. In the case of cost minimisation, the coherent risk measure would define the highest expected cost over D, while in the case of profit maximisation it would define the lowest expected profit over D. This is directly analogous to the direct definition of CVaR that we employ in Section 5.4.2.

In order to bridge between the risk neutral design game and the risky design equilibrium game, Ralph & Smeers introduce an additional agent, the system risk agent. The risk set of this agent is the intersection of all individual risk sets and is known as the system risk set. The agent's objective is based on the sum of all agent costs, or profits, as the case may be. The objective is to define the risk measure over the set of probability densities that is the system risk set. As all agents risks sets are included, the price of risk, which is the PDF that solves their optimisation, defines the payoff for the system agent. We could imagine other participants trading risk with the system agent on this basis. The system agent is risk neutral with respect to all risks in its risk set, so that in equilibrium all agents must also face that same price, or marginal view, of risk (probability density function). They can then trade as if they were risk neutral with respect to the PDF defined by the system risk agent. Unlike in the case of portfolio optimisation, the PDF that prices risk is determined by equilibrium in the risk market. Ralph & Smeers (2015) continue by generalising their result in the case of incomplete markets.

With respect to this approach, the objective functions of the agents are defined in Section 5.5. and Section 5.6. Risk market clearance is also defined in Section 5.6 and, albeit in a much narrower setting, we confirm that the most risk neutral participant sets the price of risk in this setting. To simplify discussion throughout, we define the CVaR set to be the "optimal" risk set, representing the actual selection from D, the wider risk set, that represents the worst case. Finally, in our example we assume a perfect overlap between the individual risk sets of participants so that they only differ by risk aversion. Accordingly, we do not need to explicitly introduce a system risk agent.

## *5.4 Formulation of Risk*

Our analysis follows an albeit simplified approach modelled on the stochastic endogenous equilibrium approach by Smeers & Ralph (2011). We explicitly consider the equilibrium and price determination in contract markets, and within the narrow confines of our example generate results that align with the more general theory illustrated in that seminal paper. Within that framework, we consider two

approaches for modelling risk management by individual firms. The first, which we assess as being more suitable for use in complex situations such as the investment problem, is preference based risk management in which the investor actively weighs risk and returns based on their relative preference for each. This aligns with portfolio management which we discuss in Section 5.5.2. The second is the constraint based approach to risk management. This suggests, without necessarily requiring, the involvement of external entities such as financiers in setting the risk profile of the firm. We discuss risk constraints further in Section 5.5.5.

### 5.4.1 Profit Distributions

Before we consider risk measures, we review a generic variability hierarchy as shown in Figure 33. This illustrates the different layers of variability that might exist for the purpose of discussing which modelling paradigm, for example risk or uncertainty, is suitable for variability occurring on a particular time scale.

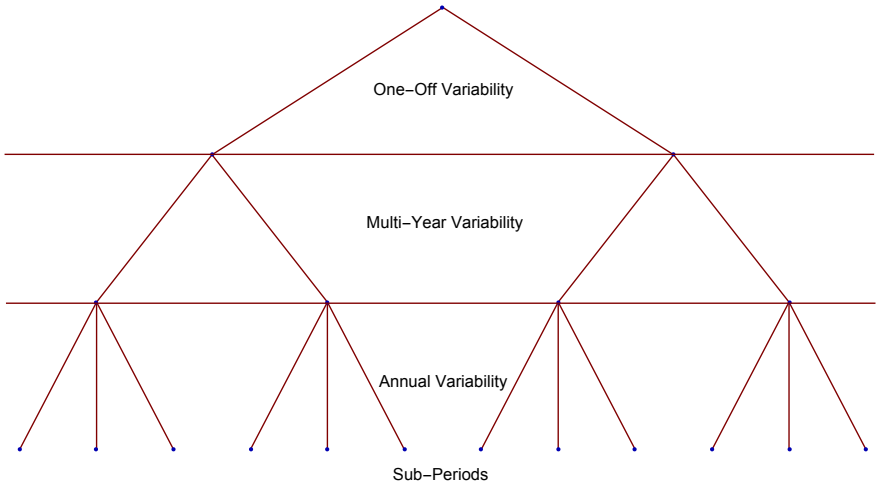



**Figure 33: Hierarchy of Variability**

Beginning with the shortest timeframes, the most fundamental sub-period variability in the investment model is represented by the LDC itself. We have implicitly assumed that we can treat load or net load as a deterministic set of proportions rather than a random variable continually being sampled. This implies a belief that the entire distribution will be observed in the timeframe of interest. The validity of treating net load as deterministic rests on how well the net LDC represents outcomes within individual periods. To ensure this, the sub-period structure must be defined with suitable granularity so that the distribution of net load within a sub-period is not unduly contaminated by seasonal influences, for example, that would lead to correlated results that would tend not to support the realisation of all load levels identified by the LDC. Sub-period modelling also enables the description of operational decisions and more frequent information revelation, such as is required to express storage and energy release decision when fuels are limited in supply. In our case, except for stochastic energy limits as discussed in Appendix 7.3, we do not model stochasticity at this level.

At the period level, which is annual in this case, we give consideration to variability in the macro system state. In our case, variability in annual hydrological and climate conditions provide

motivation for this level of variability. The process remains the same as for sub-period modelling. By identifying different scenarios, the variability within each sub-period is refined by removing sources of intra-period correlation out of the sub-period structure and placing it into the scenario structure. A scenario-based approach allows explicit specification of those correlations through a scenario tree. For example, we include scenarios for both hydrology and climate and assume that they are correlated with each other. The resulting profit distribution can be represented by a multiple level index reflecting the tree, or as we do, a distribution of flattened scenarios, indexed by s.

The longer the timeframe and more significant the variability is, the more important multiple the difference between a proportional treatment and a full expansion of the decision tree becomes as entire periods become correlated. For example, when we consider correlations that apply over a longer, such as a supra-annual climate cycle, several periods will have correlated results. A Markov chain, with an associated transition matrix as shown in (5.1) can represent the behaviour of such a climate cycle or more general evolution of the system state and by doing so capture the nature of correlation between periods:

$$\begin{pmatrix} x & 1-x \\ 1-x & x \end{pmatrix} \qquad \forall f, 0 \le x \le 1 \qquad (5.1)$$

Where $x$ is close to unity, the system state is highly correlated with the previous state, whereas when x is zero, the state switches from one state to another. When x is 0.5, the system state is independent of previous system states. In all cases the long run proportion of time spent in each state is equal, but the correlation between performance in one period with performance in the next is dependent on the value of x in this case. So while we could accurately model a proportion, that would not reflect the dynamic structure of state development.

Ultimately, we might consider a risk that corresponds to a single eventuality. Once determined, certain eventualities are likely to be permanent, or at least prevail for a significant time in comparison to the life of an investment. This also is beyond our scope but we note that outcome of variability of this nature may be inherently difficult to characterise with a distribution and therefore the situation may be better aligned with uncertainty than risk.

The first step in a mathematical formulation of risk is the definition of the distribution of outcomes. We define the profit distribution using s, an index of the flattened set of possible scenario combinations at the annual level in the decision tree in Figure 33. We assume that we can quantify the relative probability of each scenario s, and therefore we are able to treat the variability as risk rather than uncertainty. The evaluation of risk requires an assessment of the distribution of outcomes. Either preferences or constraints can be used to guide behaviour and these are stated in terms of risk measures, which define a particular characteristic in the distribution of outcomes that investors do not desire.

Although heterogeneous generators with different initial holdings of each generation technology and/or different preferences with respect to risk can be considered, we envisage G symmetric generation firms operating under competitive circumstances, indexed by g=1…G,.

Naturally generators do not "invest" in the notional shortage technology. The total annualised operating profit for generator g is given by:

$$\pi_g^{op} = \sum_{i>0} CAP_{g,i} \sum_s \mathrm{w}_s \sum_t \mathrm{w}_{s,t} \chi_{i,s,t} \qquad \forall g \qquad (5.2)$$

Where $\mathrm{w}_{s,t}$ represents the weighting of sub-period t in scenario s, and $\chi_{i,s,t}$ represents the earnings per unit of capacity of technology i in sub-period t of scenario s. After allowing for the amortisation of fixed costs, the generator's total annualised profit is given by:

$$\pi_g = \sum_{i>0} CAP_{g,i} \sum_s \mathrm{w}_s \sum_t \mathrm{w}_t \chi_{i,s,t} - \sum_{i>0} \mathrm{FC}_i CAP_{g,i} \qquad \forall g \qquad (5.3)$$

This can also be expressed as the distribution of percentage returns across the scenarios s, where the percentage return for scenario s is given by:

$$\frac{\sum_{i>0} CAP_{g,i} \sum_t \mathrm{w}_t \chi_{i,s,t}}{\sum_{i>0} \mathrm{FC}_i CAP_i} - 1 \qquad \forall g,s \qquad (5.4)$$

Figure 34 shows a hypothetical profit distribution, expressed as an annualised percentage return, across a range of scenarios for a single generation firm. The definition of profitability as an annual percentage return aligns with basic equity market valuation approaches and is independent of market or firm size. It is also invariant with respect to intra-year variations in profitability that might arise from seasonal influences, and therefore do not reflect risk.

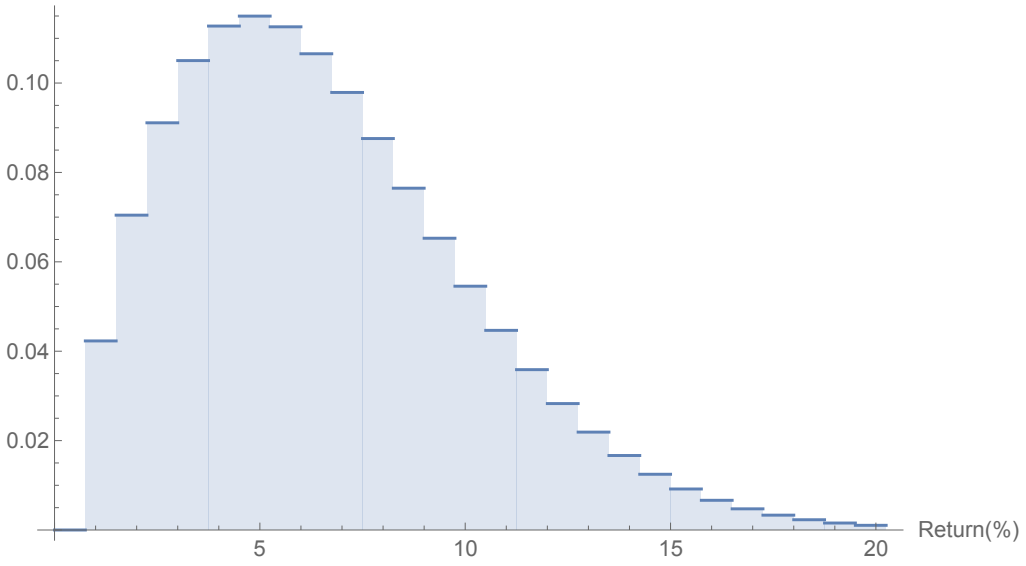



**Figure 34: Distribution of Annual Percentage Returns**

The rate of return of each technology in each individual scenario, s, is:

$$\frac{\sum_{t} w_t \chi_{i,s,t}}{FC_i} - 1 \qquad \forall i \qquad (5.5)$$

In the absence of capacity restrictions and risk aversion, the mean rate of return on investment is identical across all built technologies, reflecting a single return on investment requirement. In a standard risk neutral analysis this single rate of return is zero in equilibrium. We must be careful to note that in this context, the rate of return requirement is in excess of the risk-free rate of return, which is used purely as a measure of the time value of money in the development of the amortised fixed costs of investment. Absent investment or capacity restrictions, any deviation from the single rate of return would imply a misallocation of capital between individual technologies. To maximise portfolio returns and achieve equilibrium, investor funds would move from lower return investments towards higher return investments. Which restrictions of the form discussed in Chapter 3, the equalisation of returns across all technologies may not be possible.

Whereas the risk neutral investor is concerned only with expected returns, the risk averse investor is concerned with the distribution of returns. The mean return for the portfolio and each individual technology is identical under the assumption of investor risk neutrality. The same cannot be said under risk aversion. In this case, the distribution of returns for each individual technology varies so that ceterus paribus, risk averse investors will evaluate each technology differently, preferring an investment exhibiting less return volatility to one with greater return volatility. Investors are equally interested in the correlation between the return distributions of individual technologies. Ranking the returns of each individual technology by scenario produces a ordering of scenario outcomes that need not correspond with the ranking of the representative market portfolio across scenarios.

### 5.4.2 CVaR Calculation

We have adopted a different approach to defining CVaR, choosing instead to adopt a direct optimisation in which the conditional probabilities that we seek are the primal, as opposed to dual variables. From the optimisation, complementarity conditions are formed for the purpose of integrating with the rest of the framework. The definition of CVaR is only a subset of the conditions responsible for investment decisions. Unlike models with a purely CVaR objective, it is separate, and feeds into the overall trade-off between risk and return. The expression of CVaR as a separate optimisation, intuitively based at the risk end of the distribution, gives rise to the possibility or possible interpretation of the CVaR optimiser as a separate agent, specifically concerned with risk, that penalises investors for assuming risk. This possibility can assist in clarifying the nature of the motivation to control risk.

In accordance with the reporting structure and valuation methodology of the financial industry, our assessment of CVaR relates to the annual timescale. While this assumption is not technically necessary, it does facilitate further analogy with finance and therefore is a front-running candidate to describe how investors might financial performance in evolved electricity markets, in which generators are often publicly listed companies judged by the market. From (5.2), the annualised financial loss for a generator in a given scenario is defined as:

$$Loss_s = \sum_i \mathrm{FC_i} CAP_i - \sum_i \pi_{i,s} CAP_i \qquad \forall s \tag{5.6}$$

Where $\pi_{i,s}$ is the operating profit of technology i in scenario s:

$$\pi_{i,s} = \sum_t \mathrm{w_{s,t}} \chi_{i,s,t} \qquad \forall i,s \tag{5.7}$$

These losses, together with the weighting of each scenario, $\mathrm{w_s}$, define the distribution of losses, from which the expected loss can be determined.

$$\mathrm{E}\left[Loss\right] = \sum_i \mathrm{FC_i} CAP_i - \sum_s \mathrm{w_s} \sum_i \pi_{i,s} CAP_i \tag{5.8}$$

CVaR represents the expected value of losses greater than VaR, where VaR is defined as a tail loss where the size of the tail is defined by $\alpha^{VaR}$. It can be defined directly by selecting an alternative weighting scheme $\alpha_s$ that maximises the weighted average of losses in the tail, subject to the weight of each scenario not exceeding the objectively determined probability of each scenario and the total weight matching the pre-defined VaR level. This approach is essentially the dual perspective of the approach taken in Rockafellar & Uryasev (2000) and Ehrenmann & Smeers (2011). The formulation is tailored to the discrete scenarios we contemplate in this structure and is in that sense less general than the formulation in Rockafellar & Uryasev (2000), but it aligns more directly with our future purposes than the aforementioned formulations, which ultimately need to resort to the dual to extract the desired probabilities. To define CVaR, we use the following optimisation problem for a generator:

$$\underset{\alpha_s}{Maximise} \qquad \sum_s \alpha_s \left[ \sum_i \mathrm{FC_i} CAP_i - \sum_i \pi_{i,s} CAP_i \right] \tag{5.9}$$

Subject to:

$$\sum_s \alpha_s = \alpha^{\mathrm{VaR}} \quad : \kappa^{VaR} \tag{5.10}$$

$$0 \le \alpha_s \le \mathrm{w_s} \quad : \kappa_s^{-,+} \qquad \forall s \tag{5.11}$$

As would a separate agent tasked with assessing risk, this optimisation maximises the weighted average of losses in the tail of the loss distribution defined by $\alpha^{VaR}$. Were (5.10) to be omitted from the formulation, the optimisation would apply the maximum weighting to all positive scenario losses, which may or may not exceed the proportion of scenarios, $\alpha^{VaR}$, that we are concerned with. The inclusion of (5.10) not only prevents the optimisation from selecting all positive losses, but also may force it to select some negative losses, or low profits, in the event these outcomes fall within the portion of the loss distribution CVaR is concerned with. The final constraint guarantees that no loss can be negatively weighted, or weighted more than the actual likelihood of the loss occurring. This bounds the problem and prevents an effective arbitrage in which profitable scenarios would be assigned negative weightings to enable increased weightings for loss scenarios. When expressed as an equivalent complementarity problem, the complementarity conditions are:

$$\sum_i CAP_i\left(\pi_{i,s} - \mathrm{FC_i}\right) + \kappa^{VaR} + \kappa_s^+ \geq 0 \quad \perp \quad \alpha_s \geq 0 \qquad \forall s \tag{5.12}$$

$$\alpha^{\mathrm{VaR}} - \sum_s \alpha_s = 0 \quad \perp \quad \kappa^{VaR} \; free \tag{5.13}$$

$$\mathrm{w_s} - \alpha_s \geq 0 \quad \perp \quad \kappa_s^+ \geq 0 \qquad \forall s \tag{5.14}$$

From (5.12), where the scenario loss exceeds VaR, then $\kappa_s^+ > 0$ so that from (5.14), the weighting of that scenario is set to its maximum, $\alpha_s = \mathrm{w_s}$. Similarly, where the scenario loss is less than VaR, it must be the case that $\alpha_s = 0$. Finally, where the scenario loss is equal to VaR, it may take an intermediate value, which will be disciplined by complementarity condition (5.13), requiring the sum of assigned probabilities to equate with the level at which the VaR is defined. The combined effect of these constraints is to define VaR and the weightings, $\alpha_s$, that can be used to define CVaR. In our formulation, the distribution of scenario outcomes is discrete, which allows the possibility of $\alpha^{\mathrm{VaR}}$ coinciding precisely with the sum of relevant scenario probabilities. In this case, the boundary scenario, $\alpha_s = \mathrm{w_s}$, with $\kappa_s^+ \geq 0$ from (5.14) leaves $\kappa^{VaR}$ free to assume any value between the losses associated with the boundary scenario s, and the next more profitable/lower loss scenario. While $\kappa^{VaR}$ is free and can define a multiplicity of solutions, the value of CVaR in each of the possible solutions is unaffected as the probabilities, $\alpha_s$, are defined uniquely and identical across all solutions. By way of analogy to the definition of risk sets, and to aid the exposition, throughout the rest of this thesis we define the set of scenarios that play an active part, $\alpha_s > 0$, in the definition of CVaR as the CVaR set.

Complementarity conditions (5.12) - (5.14) enable the calculation of CVaR, the expected value of losses greater than VaR:

$$CVaR_g = \sum_s \frac{\alpha_s}{\alpha^{\mathrm{VaR}}} \left[ \sum_i CAP_{g,i}\left(\mathrm{FC_i} - \pi_{i,s}\right) \right] \tag{5.15}$$

The scaling factor $1/\alpha^{\mathrm{VaR}} = 1/\sum_s \alpha_s$ normalises the weighting scheme to ensure the weights sum to unity, thereby creating a conditional probability for the CVaR set.

## *5.5 Equilibrium with Risk Aversion*

### 5.5.1 Introduction

We begin by assuming that there are no meaningful long term contracting options available. In some markets this is precisely the case, while in others specific market structures and contract forms are able to support investment, the analysis of which is highly dependent on the specific opportunities and challenges present. In a number of cases, long term contracting options are plagued by poor specification of risks and doubts over the ultimate value of the instruments in climates where regulators and governments have incentives to intervene in the market (Boucher & Smeers, 2012), (Finon, 2008).

Nevertheless, even in an environment without long term contracting or structural options, generators still have long term risk management strategies available to them. In the case of electricity generators, the long term decisions that pertain to the management of risk are typically investment decisions and as a result of considering risk aversion the equilibrium plant mix is altered.

Investment decisions are viewed in the context of the overall portfolio of existing generation assets, so that the problem of investment under risk is essentially a portfolio optimisation problem. While it is certainly not exclusively the case, a preference based or at least preference aligned decision structure suggests decision making by upper management or the board of directors, whose remit loosely entails the long term management and strategic positioning of the firm. We therefore consider the portfolio optimisation of generation assets as an example of risk management supported by a preference based trade-off between risk and return.

In our framework, the trade-off between risk and return is assumed to be represented by a constant parameter, but in principle there could be a set of parameters, or the parameter could itself be a variable that reflects risk aversion as a function of risk size relative to firm size, for example. Although decision makers are unlikely to have formally identified the risk aversion parameter, or parameters, that describe their preferences, from the perspective of our analysis the relevant assumption is just that decision makers do have such a preference structure and that this structure describes the trade-off between risk and return. Those preferences may be revealed by actions, or implied by equity market valuation models where the firm is publicly traded.

### 5.5.2 Resolving Risk & Return

The introduction of risk and risk aversion gives each economic agent a second criteria by which to evaluate decisions and therefore we require a mechanism to balance the typically competing objectives of risk and return. In some cases this can be achieved by a dominance relationship when comparing prospective investment projects but typically a range of non-dominated options exists. Stochastic dominance is detailed in Appendix 7.6.

To strike the preferred trade-off between risk and return, we would ideally optimise a utility function objective. Utility functions are theoretical constructs designed to represent preferences. The difficulty with a utility function approach is that the function for a particular firm is possibly unknowable, but definitely unknown and likely to exhibit significant complexity, making the econometric estimation of such a function problematic. Nevertheless, there are a number of standard functional forms that exhibit "sensible" properties when used to represent utility. Once a utility function has been selected for a particular analysis, the utility function replaces the traditional profit maximisation function. For investment problems, the utility function is typically a measure of future profits, or asset value.

The approach developed by Markowitz yields efficient frontiers but there remains the question of how to choose between different Pareto-combinations of mean and variance. To resolve the issue, we are required to balance expected returns with a risk measure. Having made the decision to proceed in that direction, we must then ask which of the risk-return objectives should be used to make that choice, and which parameter(s) describe the trade-off between the twin objectives. Several risk measures could be considered. For example, in the extreme, the combination of the expected outcome

and the worst case outcome was suggested in Ziemba (2001). But more intermediate options are typical and by way of example we consider the most common enactment of this trade-off (Huang & Wu, 2008), the mean-variance objective:

$$R(X) = \left(1-\theta\right)E(X) - \theta Var(X) \tag{5.16}$$

Here $\theta$ is a parameter that describes the relative importance of expected returns and the variance of returns. The higher the value of $\theta$ the more the investor is interested in minimising risk and the less they are interested in expected returns. There is an estimation issue with respect to the determination of the parameter $\theta$. This is a difficulty present in any preference based regime where the parameter is used to identify risk aversion (Kallberg & Ziemba, 1983). In the polar cases the investor is only interested in one objective or the other but this can lead to unrealistic results as the other objective is no longer relevant, implying ambivalence to either risk or return. While this example used variance, a variety of other risk measures such as CVaR or single-sided variances can also be combined with the expected return.

Kunzi-Bay & Mayer (2006) discuss the solution of a convex combination of expected returns and a CVaR risk measure. Their approach casts the problem as a two-stage recourse problem, in which the first stage is to maximise the objective, with the risk measure defined in the recourse sub-problem. Our implementation also contemplates a firm that wishes to minimise a convex combination of expected losses and a single CVaR which is assessed at the firm, and not technological, level. As we have adopted a complementarity formulation for the purpose of achieving other objectives, the risk measure of the equivalent sub-problem is represented by the conditions defining CVaR in Section 5.4.2. In combination with expected earnings, where $0 \le \theta \le 1$, the objective is defined as follows:

$$\left(1-\theta\right)\sum_{s} \mathrm{w}_{\mathrm{s}}\left[\sum_{i} CAP_i\left(\mathrm{FC}_{\mathrm{i}}\text{-}\pi_{i,s}\right)\right] + \theta\sum_{s}\frac{\alpha_s}{\alpha^{VaR}}\left[\sum_{i} CAP_i\left(\mathrm{FC}_{\mathrm{i}} - \pi_{i,s}\right)\right] \tag{5.17}$$

Here $\theta$ represents the risk aversion of the investor. When $\theta = 1$ the investor is only interested in risk, whereas when $\theta = 0$, the investor is risk neutral. Ignoring capacity limitations, the equilibrium investment conditions consistent with this objective represent a convex combination of the investment conditions, each based on alternative scenario weighting schemes:

$$\left(1-\theta\right)\sum_{s} \mathrm{w}_{\mathrm{s}}\left[\mathrm{FC}_{\mathrm{i}} - \pi_{i,s}\right] + \theta\sum_{s}\frac{\alpha_s}{\alpha^{VaR}}\left[\mathrm{FC}_{\mathrm{i}} - \pi_{i,s}\right] \ge 0 \quad \perp \quad CAP_i \ge 0 \qquad \forall i \tag{5.18}$$

There are several possible further interpretations available. We can express (5.18) in a form that highlights the relationship between fixed costs, expected earnings and a marginal risk penalty:

$$\left(1-\theta\right)\left[\mathrm{FC}_{\mathrm{i}} - \sum_{s} \mathrm{w}_{\mathrm{s}}\pi_{i,s}\right] + \theta\left[\mathrm{FC}_{\mathrm{i}} - \sum_{s}\frac{\alpha_s}{\alpha^{VaR}}\pi_{i,s}\right] \ge 0 \quad \perp \quad CAP_i \ge 0 \qquad \forall i \tag{5.19}$$

$$\mathrm{FC}_{\mathrm{i}} - \left[\left(1-\theta\right)\sum_{s} \mathrm{w}_{\mathrm{s}}\pi_{i,s} + \theta\sum_{s}\frac{\alpha_s}{\alpha^{VaR}}\pi_{i,s}\right] \ge 0 \quad \perp \quad CAP_i \ge 0 \qquad \forall i \tag{5.20}$$

Whereas the conventional equilibrium investment constraint is based on the difference between fixed costs and the marginal profitability of additional capacity, the introduction of a CVaR risk penalty replaces marginal profitability with a convex combination of the marginal profitability and marginal impact on the CVaR constraint. The marginal CVaR penalty, which is a weighted average of the marginal profitability of technology i in the CVaR set, reflects the marginal impact on CVaR resulting from an increase in the capacity of technology i.

The difference between the CVaR adjusted marginal profitability and the risk neutral marginal profitability gives a measure of the marginal risk premium:

$$\begin{aligned}\left(1-\theta\right)\sum_{s}\mathrm{w}_{s}\pi_{i,s}+\theta\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}-\sum_{s}\mathrm{w}_{s}\pi_{i,s}\\ =\theta\left[\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}-\sum_{s}\mathrm{w}_{s}\pi_{i,s}\right]\end{aligned}\qquad \forall i \qquad (5.21)$$

Technologies can be grouped based on the relationship between the expected or risk neutral earnings and the expected profitability amongst only the CVaR set. This relationship determines the sign of the marginal risk premium, and we have three cases to consider:

- $\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}-\sum_{s}\mathrm{w}_{s}\pi_{i,s}<0$

For these technologies, expected profitability in the CVaR set is less than the risk neutral assessment of profitability, implying further investment in technology i would increase overall risk and suggesting positive correlation between the performance of these technologies and the overall portfolio. The overall correlation is only suggestive because the relevant correlation considers only that portion of the distribution used to define the risk measure. We have:

$$\mathrm{FC_i}-\left[\left(1-\theta\right)\sum_{s}\mathrm{w}_{s}\pi_{i,s}+\theta\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}\right]>\mathrm{FC_i}-\sum_{s}\mathrm{w}_{s}\pi_{i,s} \qquad (5.22)$$

Assuming the risk neutral level of capacity for each of these technologies is installed we have:

$$\mathrm{FC_i}-\left[\left(1-\theta\right)\sum_{s}\mathrm{w}_{s}\pi_{i,s}+\theta\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}\right]>0 \qquad (5.23)$$

To restore equilibrium as defined by (5.20), the installed capacity of each technology in this category must be reduced relative to their respective risk neutral levels, until expected overall earnings and/or expected CVaR set earnings increase to achieve parity with fixed costs. Ceteris paribus, the largest risk penalty, and therefore the largest deviation from the risk neutral position in this direction, will occur when the rank correlation of portfolio profitability and the profitability of technology i in the CVaR set is perfect, and positive.

- $\sum_{s}\frac{\alpha_{s}}{\alpha^{VaR}}\pi_{i,s}-\sum_{s}\mathrm{w}_{s}\pi_{i,s}>0$

For these technologies, expected profitability in the CVaR set is greater than the risk neutral assessment of profitability, implying further investment in technology i would decrease overall risk and

suggesting a negative correlation between the performance of these technologies and the overall portfolio. We have:

$$\mathrm{FC_i} - \left[ \left(1-\theta\right) \sum_s \mathrm{w_s} \pi_{i,s} + \theta \sum_s \frac{\alpha_s}{\alpha^{VaR}} \pi_{i,s} \right] < \mathrm{FC_i} - \sum_s \mathrm{w_s} \pi_{i,s} \tag{5.24}$$

Assuming the risk neutral level of capacity for each of these technologies is installed we have:

$$\mathrm{FC_i} - \left[ \left(1-\theta\right) \sum_s \mathrm{w_s} \pi_{i,s} + \theta \sum_s \frac{\alpha_s}{\alpha^{VaR}} \pi_{i,s} \right] < 0 \tag{5.25}$$

To restore equilibrium as defined by (5.20), the installed capacity of each technology in this category must be increased relative to their respective risk neutral levels, until expected overall and/or expected CVaR set earnings decrease to achieve parity with fixed costs. Ceteris paribus, the largest hedging benefit, and therefore the largest deviation from the risk neutral position in this direction, will occur when the rank correlation of portfolio profitability and the profitability of technology i in the CVaR set is perfectly negative.

- $\sum_s \frac{\alpha_s}{\alpha^{VaR}} \pi_{i,s} - \sum_s \mathrm{w_s} \pi_{i,s} = 0$

In this case risk adjusted profitability and risk neutral profitability are identical. At the margin, there is no additional risk or hedging value associated with investment in these technologies, and the risk neutral level of capacity is appropriate for these technologies.

It is important to clarify precisely what these conditions actually do, and do not, mean. The introduction of risk aversion may lead to all technologies having reduced investment, relative to the risk neutral position. As per the underlying mathematics, the conditions reflect the change in the objective function as a result of an increase in capacity of a particular technology. They do not reflect a portfolio adjustment, in which the weighting applied to one asset swings to another, as would be more typical in the context of finance. The difference is attributable to the absence of a budget constraint. In our framework there is no relativity to the investment process, and investment in individual technologies will occur whenever it can be justified. However, given a requirement to invest a fixed amount, the optimisation procedure would also ensure that this occurs where net benefits are available. In such a model, the dual on the budget constraint would normalise these marginal benefits around the portfolio return, and of necessity create relative hedges.

A further interpretation of the equilibrium investment condition is available with some minor re-arranging of (5.20):

$$\mathrm{FC_i} - \sum_s \left( \left(1-\theta\right) \mathrm{w_s} + \theta \frac{\alpha_s}{\alpha^{VaR}} \right) \pi_{i,s} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{5.26}$$

By taking the same convex combination of the scenario probabilities and CVaR set conditional probabilities, we define the equilibrium investment condition in terms of risk-adjusted probabilities. Those risk-adjusted probabilities are endogenous and defined as:

$$\omega_s = \left(1-\theta\right)\mathrm{w}_s + \theta\frac{\alpha_s}{\alpha^{VaR}} \qquad \forall s \qquad (5.27)$$

Taking advantage of this interpretation, and dividing throughout by $\mathrm{FC_i} > 0$, the investment condition is normalised in percentage terms as follows:

$$1 - \sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \qquad (5.28)$$

Risk-averse investors will equate the certainty equivalent of profits, and not the expected value of profits, with the fixed costs of investing. The difference between the certainty equivalent and the expected return is referred to as a risk premium and, ceteris paribus, the premium is greater the more risk averse the investor is, as confirmed by (5.27). The risk-adjusted probabilities are therefore specific to a particular risk attitude, and moreover, they are specific to individual firms whenever firms differ, for example, in cost structures, risk preferences or energy and capacity limitations. It is clear from (5.27) that an increase in risk aversion, as measured by an increase in $\theta$, increases the emphasis placed on poor results so that from (5.28) a reduction in investment will be necessary to regain parity with cost recovery on a risk adjusted basis. Risk aversion leads to a relative decrease in the capacity of those technologies whose profitability within the CVaR set is positively correlated with the profitability of the overall portfolio, and a relative increase in the capacity of those whose profitability within the CVaR set is negatively correlated with the profitability of the overall portfolio and therefore provide some hedging benefits.

The imposition of capacity constraints and opportunity limits will also affect the equilibration process and where those limits bind full equilibration will not be possible. Whereas in the risk neutral case, capacity limits created a disparity in returns between those technologies that were limited and those that were not, in this case there is already a disparity in the returns of individual technologies that is supported by the risk/hedging properties of those technologies. Absent capacity constraints, those differences between technologies are eliminated when we view the returns of each technology in risk adjusted terms. Under risk aversion, capacity limits prevent this equilibration and imply an equilibrium capacity mix that features disparity between not only observed returns on account of risk adjustments, but also a disparity between risk-adjusted returns as a result of capacity limitations.

### 5.5.3 Sculpting the Loss Distribution

Our framework can readily consider a number of different risk measures. The firm may be concerned about risk at a variety of different levels, and in a variety of different dimensions and may wish to control risks with more than a single constraint. For example, a firm may be relatively aggressive in assuming risk provided existential risks are avoided. Or they may be unconcerned about eventualities caused by natural occurrences, confident that the market valuation mechanisms underpinning their valuation will be relatively forgiving of poor results in circumstances that can be explained by unexpected natural variation. In general, the return distribution, representing the entire of spectrum of risk, can be sculpted by applying varying degrees of risk aversion to varying levels of risk, or to different subsets of the scenario tree. We do not consider the latter motivation here, and instead

consider a firm with a set of risk preferences that describes their attitude to a variety of different risk levels. CVaR can be assessed at a variety of different levels, which we denote by $\alpha_c^{VaR}$, so that collectively a set of CVaR preferences will effectively sculpt the profit distribution. As in Section 5.4.1, we can define a CVaR, with its own corresponding CVaR set, for each of these c levels:

$$CVaR_c = \sum_s \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \left[ \sum_i CAP_{f,i} \left( \mathrm{FC_i} - \pi_{i,s} \right) \right] \tag{5.29}$$

Each CVaR definition is represented in the model by a set of complementarity conditions that define the optimal CVaR for each risk level, c:

$$\sum_i \left[ CAP_i \left( \pi_{i,s} - \mathrm{FC}_i \right) \right] + \kappa_c^{VaR} + \kappa_{s,c}^{+} \geq 0 \quad \perp \quad \alpha_{s,c} \geq 0 \qquad \forall s,c \tag{5.30}$$

$$\alpha_c^{VaR} - \sum_s \alpha_{s,c} = 0 \quad \perp \quad \kappa_c^{VaR} \; free \qquad \forall c \tag{5.31}$$

$$w_s - \alpha_{s,c} \geq 0 \quad \perp \quad \kappa_{s,c}^{+} \geq 0 \qquad \forall s,c \tag{5.32}$$

The set of optimisations above can readily be identified as preferences held by an investor. They could equally be interpreted as relating to several entities, each with different concerns. Taking the first interpretation, the generator assigns weightings to each of the CVaR values so that a convex combination of the marginal profitability and marginal impact on CVaR forms the equilibrium investment condition:

$$\left( 1 - \sum_c \theta_c \right) \sum_s \mathrm{w_s} \left[ \mathrm{FC_i} - \pi_{i,s} \right] + \sum_c \theta_c \sum_s \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \left[ \mathrm{FC_i} - \pi_{i,s} \right] \geq 0 \perp CAP_i \geq 0 \quad \forall i \tag{5.33}$$

Note the weightings naturally sum to unity:

$$\left( 1 - \sum_c \theta_c \right) + \sum_c \theta_c = 1 \qquad \forall i \tag{5.34}$$

Following the same process as before we can develop the risk neutral probabilities associated with this preference structure. From (5.33):

$$\mathrm{FC_i} - \left[ \left( 1 - \sum_c \theta_c \right) \sum_s \mathrm{w_s} \pi_{i,s} + \sum_c \theta_c \sum_s \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \pi_{i,s} \right] \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{5.35}$$

Taking the difference between the CVaR adjusted marginal profitability and the risk neutral marginal profitability gives the following marginal risk premium for investment in technology i:

$$\begin{aligned} &\left( 1 - \sum_c \theta_c \right) \sum_s \mathrm{w_s} \pi_{i,s} + \sum_c \theta_c \sum_s \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \pi_{i,s} - \sum_s \mathrm{w_s} \pi_{i,s} \\ &\qquad = \sum_c \theta_c \left[ \sum_s \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \pi_{i,s} - \sum_s \mathrm{w_s} \pi_{i,s} \right] \end{aligned} \qquad \forall i \tag{5.36}$$

In the case where only a single CVaR penalty is considered, $\theta_c = \theta$, leaving the same expression as (5.21). The relationship between risk averse earnings and risk neutral earnings depends as before on whether technology i is correlated with the overall portfolio. In this case, the assessment is more detailed, as more CVaR constraints are involved, potentially leading to the consideration of a wider range of CVaR sets and more complex correlations in the CVaR calculation. The weighted difference between the earnings of technology i in the portfolio and each set of scenarios that correspond to each CVaR preference defines the objective so that the correlation is more nuanced than the case of a single CVaR measure. Nevertheless, it remains the case that in the presence of risk aversion, the composition of the optimal plant mix will swing towards having a relatively smaller share of those technologies whose profitability over the range of CVaR sets is generally positively correlated with the profitability of the overall portfolio, and a relatively larger share of those technologies whose profitability over the range of CVaR sets is generally negatively correlated with the profitability of the overall portfolio, and therefore provide some hedging benefits.

A further interpretation of the equilibrium investment condition is available with some minor re-arranging of (5.35):

$$\mathrm{FC_i} - \sum_s \left( \left( 1 - \sum_c \theta_c \right) \mathrm{w_s} + \sum_c \theta_c \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \right) \pi_{i,s} \geq 0 \quad \perp \quad CAP_i \geq 0 \quad \forall i \tag{5.37}$$

The risk neutral probabilities are defined as follows:

$$\omega_s = \left( 1 - \sum_c \theta_c \right) \mathrm{w_s} + \sum_c \theta_c \frac{\alpha_{s,c}}{\alpha_c^{VaR}} \qquad \forall s \tag{5.38}$$

We can express (5.35) in terms of rates of return and risk neutral probabilities as follows:

$$1 - \sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \quad \perp \quad CAP_i \geq 0 \quad \forall i > 0 \tag{5.39}$$

### 5.5.4 Equilibration of Investment

When we introduce risk aversion in the form we have, the investor is concerned about a convex combination of two or more return distributions; the distribution of returns across all scenarios, and, for each of the CVaR measures considered, the conditional distribution of returns across the relevant CVaR sets. The derivation in Section 5.5.2 illustrates how the basic convex combination of earnings can be rearranged to an alternatively weighted form. For each scenario, the scenario PDC and, by extension, profitability is combined using alternative endogenously determined risk adjusted weights based on investor preferences.

We wish to re-examine the marginal benefit function for an individual technologies in order to observe the implications of risk aversion. The introduction of scenarios, per se, does not impact the basic form of the marginal benefit function other than to further fragment the piecewise structure. Accordingly, as utilisation levels are endogenous in our framework, the option value or weighted profitability of an incremental unit of capacity also shrinks in a piecewise linear fashion. The

additional fragmentation results from the increased potential for incremental capacity to cause a discrete change to the maximum marginal cost in an increasing number of scenarios and sub-periods.

To avoid unnecessary complication, we assume the LDC is linear throughout this section, and the merit order is constant across sub-periods and scenarios. Where this is not the case, there may be variations in the rate of change in utilisation that have direct implications for the marginal benefit function and, depending on the shape of the LDC, the result will be a non-monotonic change in the gradient of the marginal benefit function. To be clear, it is the gradient of that function that adjusts non-monotonically, not the benefit function itself, which remains (weakly) monotonically decreasing.

Our assumption is only for illustrative purposes, so that we may clarify the two separate effects that are introduced when risk aversion is considered.

*Stable CVaR Sets*

The assumption of a stable CVaR set means that the set of scenarios that define CVaR and, more specifically, the set of weightings that define the CVaR measure, is constant. In a local sense this assumption is generally true. When considering more significant adjustments, increasing the capacity of a particular technology may make the firms portfolio susceptible to new risks, while potentially eliminating others, leading to a re-assessment of which scenarios pose the greater risk to the firm. Ultimately the assumption is implausible as eventually the addition of enough capacity of a particular technology will completely dominate the expected portfolio performance, and radically adjust the CVaR set. Nevertheless, we now consider the adjustment of the marginal benefit function using two examples for which CVaR scenarios are stable.

Example 1: Load Risk

In the specific case of load risk we can anticipate the CVaR set. Depending on the contract position of the firm, either high load or low load will be problematic, and in this example we assume that low load levels are the risk the firm is focussing on. Figure 35 shows the PDC based on objective weightings, and a similar construction using conditional probabilities or weightings based only on the CVaR set.

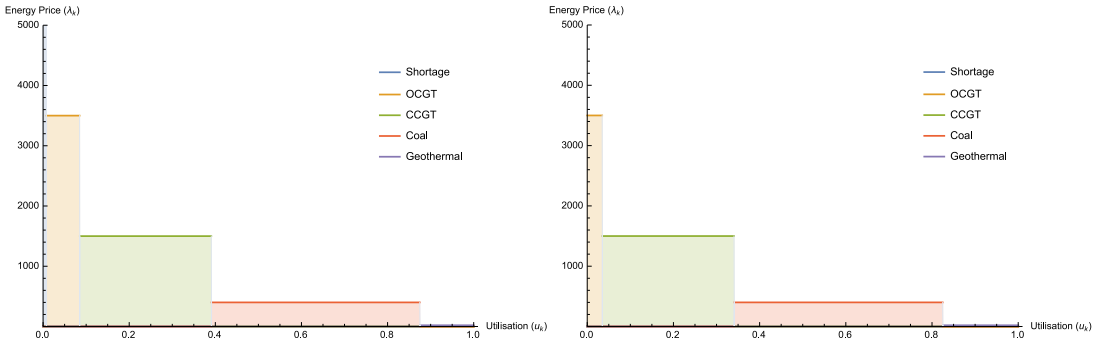



**Figure 35: Expected and CVaR PDC's with Load Risk**

As shown, the conditional PDC for the CVaR set, shown on the right, has shifted left on account of available capacity being higher than these scenarios alone would support. In relative terms, high price periods are truncated and low price periods are extended. Figure 36 shows the marginal benefit

function implied by a weighted average of these PDC's alongside the marginal benefit function for a risk neutral investor.

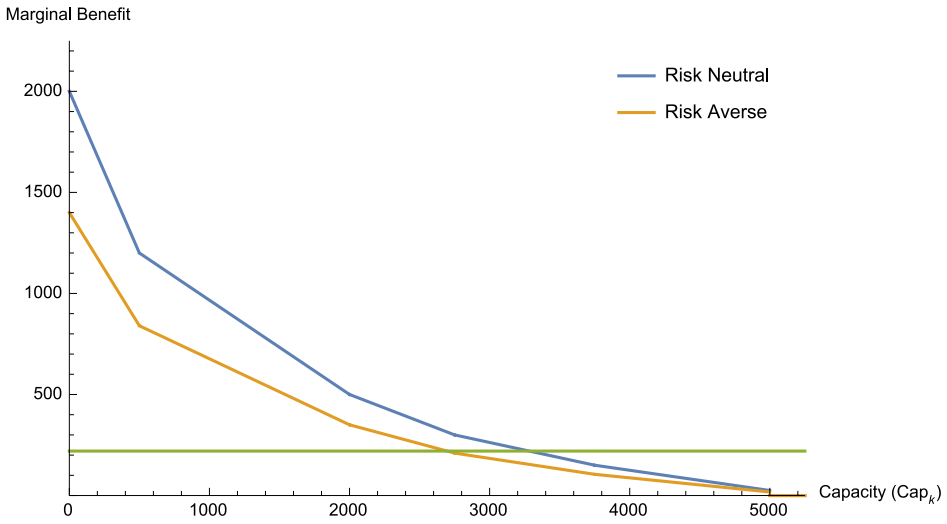



**Figure 36: Marginal Benefit Function with Load Risk**

Naturally, the influence of CVaR is to reduce the optimal level of capacity. The nature of risk aversion is to lower the marginal benefit of investment and decrease optimal investment. As capacity is added, then according to both the expected or CVaR measure, other technologies are displaced and the marginal benefit of additional capacity decreases at a rate determined by the marginal profitability of the peak technology. As result the functional form of the marginal benefit function remains piecewise linear, with the rate of decrease based on a combination of expected and CVaR scenario returns weighted by the risk aversion coefficient. When sufficient capacity is accumulated in either the overall or CVaR set to change the marginal technology, then the slope of the marginal benefit function will also change. The change is discrete and is measurable by the decrease in the market price during peaking periods under whichever scenario set initiates the change. Under perfect competition that difference will also equal the difference in cost price between the old and the new peaking technology. Our approach, using endogenous utilisation levels based on optimal trade-offs will display this effect, whereas the conventional optimisation formulation with uplift components will distort this effect somewhat. As before, the marginal benefit function is derived by varying the capacity of a single technology from zero upwards while maintaining fixed capacity of other technologies. In the case of risk neutrality, these capacity levels could readily be interpreted as equilibrium capacity levels but, when comparing risk neutral marginal benefit functions and risk averse marginal benefit functions, the issue is more complicated as the relevant equilibrium capacity level for other technologies is different in each case.

Our presentation treats the PDC's as if they were merged or averaged, with a relative weighting defined by the risk aversion parameter. This leads to a minor misunderstanding of the form of the marginal benefit function. Separate consideration of each PDC would ensure recognition of the fact that the expected marginal benefit available and conditional marginal benefit applicable to the CVaR set adjusts at a different pace and individually experience discrete changes in the rate of the

adjustment at different load levels. In contrast, the combined viewpoint effectively re-orders the returns so that equal returns in the overall and CVaR scenarios are grouped together even though they correspond to different capacity levels of the technology under consideration. The final equilibrium remains a weighted average of the returns and so is unaffected by the manner of the decomposition used.

Example 2: Fuel Price Risk

We consider the example of a non-dominant technology whose profitability within the CVaR set is, at least initially, negatively correlated with the profitability of the portfolio, so that it provides some hedging benefits and has a higher marginal benefit that its spot market returns suggest. For simplicity we assume that there is only one such technology. In this example, the CVaR set correspond to scenarios in which there are large increases in the cost price of fuel for the dominant technology in the portfolio. In those scenarios, this hedging technology supplants the technology with significant fuel risk in the merit order and has a higher utilisation and profitability than is normally the case.

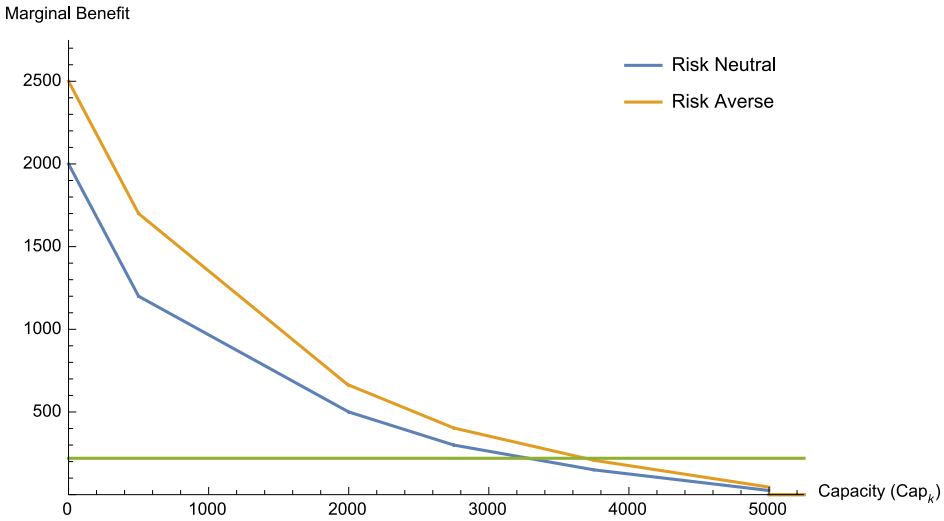



**Figure 37: Marginal Benefit Function with Fuel Price Risk**

Figure 37 shows the technology has a higher marginal benefit under risk aversion on account of its hedging properties. This is a direct result of the risk weighted average of profitability being higher than the objectively weighted profitability. The adjustment of the marginal benefit function is identical under risk aversion and risk neutrality until we reach capacity levels at which the technology subject to fuel risk begins to be supplanted. The slope of the marginal benefit function diverges at this point, becoming flatter, and remains flatter until either sufficient capacity has been built to eliminate the risky technology entirely from the plant mix, or the optimal level of installed capacity of the hedging technology is reached.

*Unstable CVaR Sets*

There is likely to be interaction between different risks, or risks that correlate more strongly with specific technologies, and in so in general we need to consider each scenario separately as they cannot

be combined using an objective scenario weighting or probability. Therefore, at least in the form we have used, the introduction of CVaR as a risk measure creates a further complication in the derivation of the marginal benefit of investment function. In principle, the CVaR optimisation chooses the best description of risk from amongst all possible combinations that satisfy the significance level constraint. Each potential combination of scenarios satisfying the constraint on significance represents a feasible solution to the CVaR optimisation problem described in Section 5.4.2. Figure 38 presents a simplified but indicative view of possible CVaR sets and the value of CVaR in each as capacity is adjusted. Each possible CVaR set is denoted a CVaR combination. Precisely drawn, the functions shown should be piecewise linear rather than linear, for the same reasons as discussed before, however this complication is not significant in this discussion.

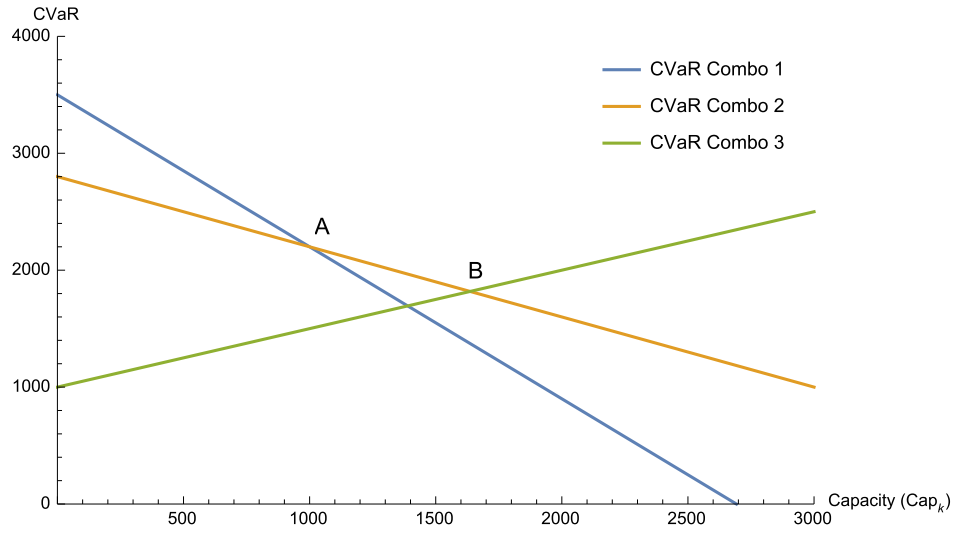



**Figure 38: CVaR Set Combinations and Capacity Choice**

For some CVaR combinations, additional capacity increase the associated conditional expected profit of that combination of scenarios, while for others it decreases. The actual CVaR measure for a given capacity choice is represented by the supremum of all such functions, which defines the worst performing, or highest loss, CVaR combination at each level of capacity and satisfies the complementarity conditions (5.30)-(5.32). In the case shown, the technology under consideration initially provides a hedge against those outcomes that are worst for the overall portfolio.

As capacity increases from zero towards A, conditional losses for CVaR combination 2 are decreasing at a lower rate than the actual CVaR set, defined by CVaR combination 1. At the same time losses in CVaR combination 3 are increasing. In general, there could be any combination of CVaR combinations, adjusting relative to the actual CVaR set. At A, the CVaR set is on the verge of changing from CVaR combination 1 to CVaR combination 2 as at this point CVaR combination 2 becomes the marginal combination and defines the supremum of all CVaR scenario combinations. This change represents the focus of risk changing from one scenario to another. Moving beyond A represents a discrete change in the scenario weightings defined by the complementarity conditions (5.30)-(5.32), or in linear programming terms, movement to a new basis as one scenario is removed and another added to the optimal set. As there is no restriction on the number of scenarios that may

participate in the definition of the CVaR set, then at A we have a range of solutions to the CVaR defining problem. These represent all of the feasible convex combinations of the weightings associated with scenario entering focus and the scenario leaving focus. The two basic solutions corresponding to the boundary of that set of solutions are equal in terms of the measure they define, so although the conditional PDC describing the CVaR set is discretely adjusted by the revision of scenario weights, the marginal benefit of investment remains the same, implying continuity of that function. As the functions above are monotonic, it must also be the case at each intersection that the gradient of the conditional profitability function of the CVaR combination entering the CVaR definition is greater (less negative) than that for the existing CVaR set, so the slope of the marginal benefit function changes discretely at this capacity level.

Eventually, as shown in Figure 38 at the point B, the level of capacity for this technology reaches a level at which the overall performance of the portfolio becomes correlated with its own performance. At this point, additional investment will worsen CVaR so that the technology no longer provides a hedge, and becomes risk enhancing. We discuss this issue further in Section 5.6 in the context of contracting.

*Conventional Optimisation Formulations*

As discussed in Section 1.6.2, when adopting the conventional optimisation formulation, the adjustment in marginal benefits is restricted to price adjustment alone as utilisation levels are fixed. When investment results in a particular technology being marginal and at full capacity, the traditional approach relies on non-competitive prices, including a capacity uplift or cost recovery component to support investment.

When risk is introduced the same problem remains. By virtue of the risk premium introduced, whether directly or implicitly through the use of risk adjusted weightings, the capacity uplift that is required by the investment constraint is increased. In the risk neutral model, spot market prices were defined by capacity cost recovery post investment. With the introduction of risk, we have an even more inconsistent situation in which spot market prices are also, to some degree, determined by the risk premium on investment. Investors would have to assume that once built, the market operator would respect their own assessment of the required risk adjustment in spot market pricing, or if viewed dynamically, that the spot market pricing would support the required risk adjustment of future investors.

Given the components of the spot market price now include a cost recovery component and a risk premium on that, the question for investors becomes which part of the observed price is which. The situation is further muddied when we consider that the risk of other technologies directly influences the spot market pricing of all other technologies as other technologies capture the benefits of risk premiums associated with other technologies while inframarginal.

While unpalatable in a theoretical model such as this, this type of logic is factored into the determination of price caps in Australia, for example. Where that occurs though, we prefer to model that explicitly to ensure the form of consideration is correctly anticipated rather than being indiscriminately applied to the PDC in general.

### 5.5.5 Risk Constraints

When viewed from the perspective of the preference system that they imply, risk constraints appear to have weak conceptual foundations. Risk constraints imply that no amount of additional profit could persuade the firm to assume incrementally more risk than allowed by the risk constraint. This is unlikely to be true, but in a wider context, risk constraints may not represent preferences alone. They may arise as a result the transaction costs of managing risk, including the need to benchmark risk, as well as the direct costs of managing risk. That aside, constraints are the conceptual basis of the original approach of Markowitz, although that form of portfolio optimisation envisaged more than a single interpretation of the constraint. Instead, that approach involved the parametric consideration of either risk or return constraints, while optimising the other to produce a Pareto-efficient frontier from which investors could choose.

The Markowitz approach can be extended with addition of preferences to develop a solution on the Pareto efficient risk-return frontier. Krokhmal et al (2002) were first to integrate CVaR constraints into optimisation problems. Their approach is based on forming an estimate of the efficient frontier, perhaps of a linear form, for consideration in the objective. Others such as Fabian (2008) also consider the problem of integrating risk constraints into a stochastic programming framework, focussing on algorithmic issues surrounding the development of an approximately efficient frontier from which a representation of CVaR is established. Both of these approaches are amenable to the definition of several CVaR constraints. These decompositions provide a useful conceptual basis with which to consider the means by which investment choices are made from among the alternatives available. However, in an equilibrium model the position of those frontiers is endogenous, which adds some complication to the issue.

Our implementation proceeds on the same basis as before. CVaR preferences are formulated utilising the definition of CVaR provided by the complementarity conditions in Section 5.4.2. In order to comply with external requirements such as financing covenants, the firm must operate within limits and therefore must implement a constraint of the following form that limits CVaR to $\mathrm{CVaR}^{+}$, with the corresponding dual variable $\kappa^{CON} \geq 0$ reflecting the marginal benefit available from a relaxation of the constraint:

$$\mathrm{CVaR}^{+} - \sum_{s} \frac{\alpha_s}{\alpha^{VaR}} \left[ \sum_{i} CAP_{f,i} \left( \mathrm{FC}_{\mathrm{i}} - \pi_{i,s} \right) \right] \geq 0 \quad \perp \quad \kappa^{CON} \geq 0 \tag{5.40}$$

As a result, the equilibrium investment condition is modified to:

$$\sum_{s} \mathrm{w}_{\mathrm{s}} \left[ \mathrm{FC}_{\mathrm{i}} - \pi_{i,s} \right] + \kappa^{CON} \sum_{s} \frac{\alpha_s}{\alpha^{VaR}} \left[ \mathrm{FC}_{\mathrm{i}} - \pi_{i,s} \right] \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{5.41}$$

$$\mathrm{FC}_{\mathrm{i}} - \left[ \sum_{s} \mathrm{w}_{\mathrm{s}} \pi_{i,s} - \kappa^{CON} \left( \mathrm{FC}_{\mathrm{i}} - \sum_{s} \frac{\alpha_s}{\alpha^{VaR}} \pi_{i,s} \right) \right] \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \tag{5.42}$$

Where the CVaR constraint is slack, then from (5.40), we have $\kappa^{CON}=0$ so that the risk adjusted and risk neutral equilibrium capacity conditions coincide. Where the CVaR constraint is binding, we have a risk-based adjustment to the profitability of technology i:

$$-\kappa^{CON}\left(\mathrm{FC_i}-\sum_s \frac{\alpha_s}{\alpha^{VaR}}\pi_{i,s}\right) \qquad \forall i \qquad (5.43)$$

As before, the sign and magnitude of the risk adjustment is determined by the correlation between the profitability of technology i, and the overall portfolio. To the extent that technology i is correlated to the overall portfolio, then technology will incur losses in those scenarios that define CVaR. $\mathrm{FC_i}-\sum_s \frac{\alpha_s}{\alpha^{VaR}}\pi_{i,s}$ will be greater for technology i than for those other technologies where the correlation is less positive. As $\kappa^{CON}>0$, technology i attracts a greater risk penalty, thereby requiring a reduction in the capacity of technology i relative to the risk neutral capacity that would be selected. The reverse holds for those technologies whose profitability is negatively correlated to the overall portfolio. They will have greater installed capacity relative to their risk neutral install capacity. Finally, where $\mathrm{FC_i}-\sum_s \frac{\alpha_s}{\alpha^{VaR}}\pi_{i,s}=0$, technology i attracts no risk premium or discount.

We are also able to view the equilibrium investment constraint in terms of risk neutral probabilities. From (5.42) we have:

$$\left(1+\kappa^{CON}\right)\mathrm{FC_i}-\sum_s\left(\mathrm{w_s}+\kappa^{CON}\frac{\alpha_s}{\alpha^{VaR}}\right)\pi_{i,s}\geq 0 \quad \perp \quad CAP_i\geq 0 \qquad \forall i \qquad (5.44)$$

Normalising by a factor of $1/\left(1+\kappa^{CON}\right)$ gives:

$$\mathrm{FC_i}-\sum_s\left(\frac{\mathrm{w_s}}{1+\kappa^{CON}}+\frac{\kappa^{CON}}{1+\kappa^{CON}}\frac{\alpha_s}{\alpha^{VaR}}\right)\pi_{i,s}\geq 0 \quad \perp \quad CAP_i\geq 0 \qquad \forall i \qquad (5.45)$$

We confirm that the weights sum to unity:

$$\begin{aligned}\sum_s\left(\frac{\mathrm{w_s}}{1+\kappa^{CON}}+\frac{\kappa^{CON}}{1+\kappa^{CON}}\frac{\alpha_s}{\alpha^{VaR}}\right) &= \frac{1}{1+\kappa^{CON}}\sum_s \mathrm{w_s}+\frac{\kappa^{CON}}{1+\kappa^{CON}}\sum_s\frac{\alpha_s}{\alpha^{VaR}} \\ &= \frac{1}{1+\kappa^{CON}}+\frac{\kappa^{CON}}{1+\kappa^{CON}}=1\end{aligned} \qquad (5.46)$$

The risk constraint also coincides with the convex combination interpretation noted earlier. We define the risk neutral probability as

$$\omega_s=\frac{1}{1+\kappa^{CON}}\mathrm{w_s}+\frac{\kappa^{CON}}{1+\kappa^{CON}}\frac{\alpha_s}{\alpha^{VaR}} \qquad \forall s \qquad (5.47)$$

Risk adjustments are implied by preferences and/or constraints, so that for every binding risk constraint there is an implied risk preference, and vice versa. In this case, the preference and constraint based risk measures are identical when $\theta = \frac{1}{1+\kappa^{CON}}$.

Substituting the risk neutral probabilities and dividing all terms by $\mathrm{FC_i}$, we arrive at the same equilibrium investment condition as in (5.28), albeit with a different definition of the endogenous risk neutral probabilities:

$$1 - \sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \perp CAP_i \geq 0 \qquad \forall i > 0 \qquad (5.48)$$

From the risk measure we have developed, we are able to define an optimal portfolio frontier in risk-return space by varying the CVaR constraint. To cast the problem in more identifiable terms, consistent with the usual presentation, we normalise the constraint and the model output to percentage returns. To do this we rely on the positive homogeneity of CVaR as a coherent risk measure. If investors wish to elicit some preference information, parametrically varying the CVaR constraint will allow them to construct an efficient risk return frontier so that the sensitivity of profit to risk taking can be understood in terms that coincide with standard business practice.

## *5.6 Contracting*

### 5.6.1 Introduction

The primary source of risk management in electricity markets is through contracting. Contracting redefines and reassigns risk between participants. That said, risk, or contract, markets are unlikely to be complete. Contracts designed to eliminate long-term project risk, either for a major consumer, for the investor, or the investor's financier, are often hard to find in many markets (Deng & Oren, 2006). These sorts of contract may be required to secure funding for a significant capital project and in many countries this surety may have to be provided by the government. In the absence of such sureties, vertical integration represents a de facto strategy where markets are not performing well or are thin, and where risk premiums are high (Meade & O'Connor, 2009). The nature of the availability of these contracts and/or the possibility of vertical integration is specific to individual markets and we do not focus on these in this research.

Although there are typically a range of contracting options, transaction costs prevent the identification, development and trading of a definitive range of contracts to cover every conceivable risk. Contracts are typically written on the basis of outcomes rather than the underlying reasons for a particular market eventuality, although exotic options that are directly related to specific causal eventualities, such as temperature or weather, have been created (Lee & Oren, 2009) and, at least in principle, these offer the promise of inter-sector trading of risk between entities who share no common outputs. Within a sector, a contract focussed on outcomes rather than causes is not necessarily problematic, as investors ultimately care about the distribution of market outcomes and not the

underlying reason for them. Problems can arise with standardised contracts when they fail to account for timeframes and for certain types of risk, or when parties lack a counter party, or the counter party is asymmetric and has alternative risk management tools available to them in the medium or longer term.

There are many different contracts that might be applied to a given risk position, but a few are particularly relevant to electricity markets, in which a core of standard contract forms are commonly traded, in addition to a much smaller cohort of specialised contract forms offering protection against specific types of risk. Among the basic contractual forms, we see some that are applicable to different technology types. For example, a theoretically fully reliable base load plant could significantly reduce market risk if it were to sign a CFD for its entire output. Similarly, a peaker could significantly reduce market risk by selling an option on its output with a strike price equal to its marginal cost. Such a combination would leave the owner indifferent as to whether the plant operated or not, as the revenue would come from subscription to the option. Below is a very brief summary of common contract forms:

- An option gives the right but not the obligation to buy or sell (call or put) an underlying asset at a pre-determined price.
- Forward Contracts specify a price and quantity at a specific future date, time and locations
- Futures Contracts are similar to forward contracts but, by virtue of being exchanged traded and therefore standardised, they contain some basis risk. Forward markets are promoted as an appropriate measure to improve electricity market design (Cramton, 2010)
- Swaps/Contracts for Difference allow two parties to agree a price, with one party compensating the other depending on whether the market price is above or below the agreed price. A swap can be thought of as a strip of forwards.
- FTR's are designed to allow participants to hedge locational risks that arise from the nature of the transmission system. While transmission losses cannot be hedged effectively, FTR or similar markets do allow hedging of the price differentials that arise from transmission constraints, and provide some protection from network related gaming.

A more detailed review of commonly traded contract forms is presented in Deng & Oren (2006), and Eydeland & Geman (1999). More exotic contracts, many of which are compound forms of basic contractual arrangements, are described in a review by Chase Bank (1992).

The interaction between contracting and spot market behaviour is often cast in a multiple stage context in which the contract market is resolved in advance of the spot market (Shanbhag et al., 2011), (Ralph & Smeers, 2006) & (Yao, Adler, & Oren, 2008). Others, including Batstone (2003), consider an opposing interaction, where spot market outcomes influence contract prices. Without a supporting explanation of entry deterrence or, alternatively, invoking the basic assumption that entry is restricted for some reason, the entry mechanism of perfect competition will apply equally to either market, leaving the risk adjusted return from each market equal in equilibrium. Contracts may be entered continuously and/or on a rolling basis, creating a complex information structure in which the contract term and information structure is significant along with the ability of the firm to dynamically hedge risk operationally. To simplify, we assume that contracts are entered into for a term equal to a single period, and this is done before particular hydrological, or other significant conditions, are realised.

Finally, contracts themselves are subject to their own uncertainty and/or risk, and require considerable prudential oversight if parties to a contract are to be assured, inasmuch as it is reasonably practicable to do so, that the contract can be satisfied. In what follows we ignore this aspect of contracting.

### 5.6.2 Forward Contracts

*Introduction*

Although the contracting landscape is rich, we confine ourselves to the analysis of forward contracts, these being one of the predominant contract structures in electricity markets (Deng & Oren, 2006). Forwards are financial contracts that can be defined in a variety of ways, over a variety of timeframes and groupings, enabling a portfolio of forwards to be constructed to reasonably match a certain load profile. We focus on a single simple contract designed to smooth variations in prices for both consumers and generators over an annual period, in which each party agrees to transact a quantity, FWD, at a price, $\lambda^C$ , both of which we initially assume are fixed. This represents a strip of forwards and is known as a "swap". It comprises a series of individual forward contracts, for which the reference price is the average over the period concerned. We do not consider discounting as the contract is assumed to have zero value in prospect and therefore requires no upfront payment.

*Contract Payout Mechanics*

From the perspective of the generator selling FWD units of the contract in a particular scenario, the pay-out in a particular scenario can be expressed as:

$$\begin{aligned} Contract\ Payout_s &= \mathrm{FWD}\sum_t \frac{\mathrm{w_{s,t}}}{\mathrm{w_s}}\sum_{r>0}\left(\lambda^C-\lambda_{r,s,t}\right)\left(u_{r,s,t}-u_{r-1,s,t}\right) \\ &= \mathrm{FWD}\sum_t \frac{\mathrm{w_{s,t}}}{\mathrm{w_s}}\left[\lambda^C-\sum_{r>0}\lambda_{r,s,t}\left(u_{r,s,t}-u_{r-1,s,t}\right)\right] \quad \forall s \\ &= \mathrm{FWD}\left[\lambda^C-\lambda_s^{TWAP}\right] \end{aligned} \tag{5.49}$$

Whenever the time weighted average market prices in scenario s exceeds the contract price, generators compensate contract purchasers according to the difference. Therefore, the firm's loss function in a given scenario is:

$$\sum_i CAP_i\left(\mathrm{FC_i}-\pi_{i,s}\right)-\mathrm{FWD}\left(\lambda^C-\lambda_s^{TWAP}\right) \qquad \forall s \tag{5.50}$$

Taking expectation across all scenarios we have:

$$\sum_s\sum_i \mathrm{w_s}CAP_i\left(\mathrm{FC_i}-\pi_{i,s}\right)-\sum_s \mathrm{w_s}\mathrm{FWD}\left(\lambda^C-\lambda_s^{TWAP}\right) \qquad \forall s \tag{5.51}$$

Re-arranging (7.46) gives a more intuitive statement of the objective function, defined in terms of losses, in the form of capacity costs less the sum of operating and contract trading profits:

$$\sum_i \mathrm{FC_i}CAP_i-\left(\sum_{i,s}\mathrm{w_s}\pi_{i,s}CAP_i+\mathrm{FWD}\left(\lambda^C-\lambda^{TWAP}\right)\right) \tag{5.52}$$

We draw the attention of the reader to the distinction between $\lambda_s^{TWAP}$, the time weighted average price within a scenario and $\lambda^{TWAP}$, the time weighted average price across all scenarios and note that in the same terms as (7.47), a base-load plant operating all the time at full capacity will earn spot market income of $\lambda^{TWAP} CAP_i$. Therefore, the contract price precisely covers fixed and operating costs. This well-known equivalence between the economics of base-load technologies and the contractual structure of a CFD holds only in risk neutral terms as risk adjusted scenario weightings are endogenous.

*Defining CVaR with Contracting*

Given a particular investment plan, a CVaR optimisation calculates the CVaR measure at the required level, $\alpha^{\text{VaR}}$, effectively defining the worst expectation for that portion of the profit distribution under consideration. Including contracts, the CVaR defined by our optimisation is:

$$CVaR = \sum_s \frac{\alpha_s}{\alpha^{VaR}} \left[ \sum_i CAP_i \left( \text{FC}_\text{i} - \pi_{i,s} \right) - FWD \left( \lambda^C - \lambda_s^{TWAP} \right) \right] \tag{5.53}$$

Following the same procedure used earlier, we define CVaR at $\alpha^{VaR}$ level as an alternatively weighted profit using the following optimisation.

$$\underset{\alpha_s}{Maximise} \quad \sum_s \frac{\alpha_s}{\alpha^{VaR}} \left[ \sum_i CAP_i \left( \text{FC}_\text{i} - \pi_{i,s} \right) - FWD \left( \lambda^C - \lambda_s^{TWAP} \right) \right] \tag{5.54}$$

Subject to:

$$\sum_s \alpha_s = \alpha^{\text{VaR}} \quad : \kappa^{VaR} \tag{5.55}$$

$$0 \le \alpha_s \le \text{w}_\text{s} \quad : \kappa_s^{-,+} \qquad \forall s \tag{5.56}$$

The complementarity conditions describing the solution to this optimisation are:

$$\sum_i CAP_i \left( \text{FC}_\text{i} - \pi_{i,s} \right) - FWD \left( \lambda^C - \lambda_s^{TWAP} \right) + \kappa^{VaR} + \kappa_s^+ \ge 0 \quad \perp \quad \alpha_s \ge 0 \qquad \forall s \tag{5.57}$$

$$\alpha^{\text{VaR}} - \sum_s \alpha_s = 0 \quad \perp \quad \kappa^{VaR} \; free \tag{5.58}$$

$$\text{w}_\text{s} - \alpha_s \ge 0 \quad \perp \quad \kappa_s^+ \ge 0 \qquad \forall s \tag{5.59}$$

### 5.6.3 Contract Markets

*Generators*

Typically, optimal contracting or portfolio construction more generally are defined in terms of fixed prices for contracts, or financial assets. The optimisation of the portfolio is therefore based on the choice of quantities. In the spirit of Ralph and Smeers (2011), we consider contracting in a framework in which the price and quantity of contracts traded are endogenous. We consider two broad types of risk: quantity risk and price risk. These could be proxies for underlying risks but suffice for this exercise. The static position for generators was discussed in the last section.

We begin by reintroducing the risk-neutral generator's loss function, and consider the contract quantity as a variable rather than a constant, although still with a fixed contract price:

$$\sum_i CAP_i\left(\mathrm{FC_i} - \pi_{i,s}\right) - FWD^g\left(\lambda^C - \lambda_s^{TWAP}\right) \tag{5.60}$$

The complementarity condition governing the equilibrium contract position for the generator is:

$$\lambda^{TWAP} - \lambda^C \geq 0 \qquad \perp \quad FWD^g \geq 0 \tag{5.61}$$

In the absence of risk aversion, where $\lambda^{TWAP}$, the time weighted average price, exceeds the contract value price, a generator will not sell any forward contracts. When the contract price is greater than $\lambda^{TWAP}$, full contracting will occur. We consider a risk-averse generator, whose optimal contracting condition is:

$$\theta^g\left(\lambda^{TWAP} - \lambda^C\right) + \left(1-\theta^g\right)\sum_s \frac{\alpha_s^g}{\alpha^{VaR,g}}\left(\lambda^{TWAP} - \lambda^C\right) \geq 0 \quad \perp \quad FWD^g \geq 0 \tag{5.62}$$

Which can be more conveniently stated as:

$$\left[\theta^g \lambda^{TWAP} + \left(1-\theta^g\right)\sum_s \frac{\alpha_s^g}{\alpha^{VaR,g}}\lambda_s^{TWAP}\right] - \lambda^C \geq 0 \quad \perp \quad FWD^g \geq 0 \tag{5.63}$$

Barring any restriction on contracting, risk averse generators will continue to sell forward contracts until:

$$\lambda^C = \theta^g \lambda^{TWAP} + \left(1-\theta^g\right)\sum_s \frac{\alpha_s^g}{\alpha^{VaR,g}}\lambda_s^{TWAP} \tag{5.64}$$

This represents a convex combination of the time-weighted average over all scenarios and those over the generators CVaR set. We can state the acceptable (negative) premium for generators selling contracts as:

$$\left(1-\theta^g\right)\left[\sum_s \frac{\alpha_s^g}{\alpha^{VaR,g}}\lambda_s^{TWAP} - \lambda^{TWAP}\right] \tag{5.65}$$

The premium is defined relative to the fair value. With heterogeneous participants, the premium will different for each participant, as each will face the same market price of contracts. When the CVaR set is constant the level of contracts, while lowering CVaR (and VaR) and improving the firms objective function, does not alter the marginal benefit of investment. When these scenario sets do change, the marginal benefit also changes. Starting from the point of zero contracting, increases in the contract position eventually lead to reduction in the discount generators are prepared to accept. Eventually higher contract levels reduce the discount generators are willing to offer to zero. At contracting levels beyond this, generators demand a premium for accepting risk. At these contract levels, the contract position becomes the dominant position in the portfolio and is effectively hedged by physical capacity rather than the other way around.

When risk markets are not complete and the contractual form of the contract does not align perfectly with the risks faced by the firm the CVaR set will not be stable. For example, where a firm faces not only price risk, but also some quantity risk, and the CVaR set comprises scenarios that include a mixture of outcomes associated with these individual risks, then an increase in the contract quantity will asymmetrically impact the scenarios and cause a reassessment of the scenarios that comprise the CVaR set. In contrast, where the firm's risk is entirely price related, the supply curve for contracts will only consist of two tranches; one for when under contracted and one for when over-contracted.

Figure 39 shows the nature of contract supply when the CVaR set is not stable. The marginal benefit of contracting is piecewise constant, so that throughout each contract range in which the CVaR set is constant the price is unchanged.

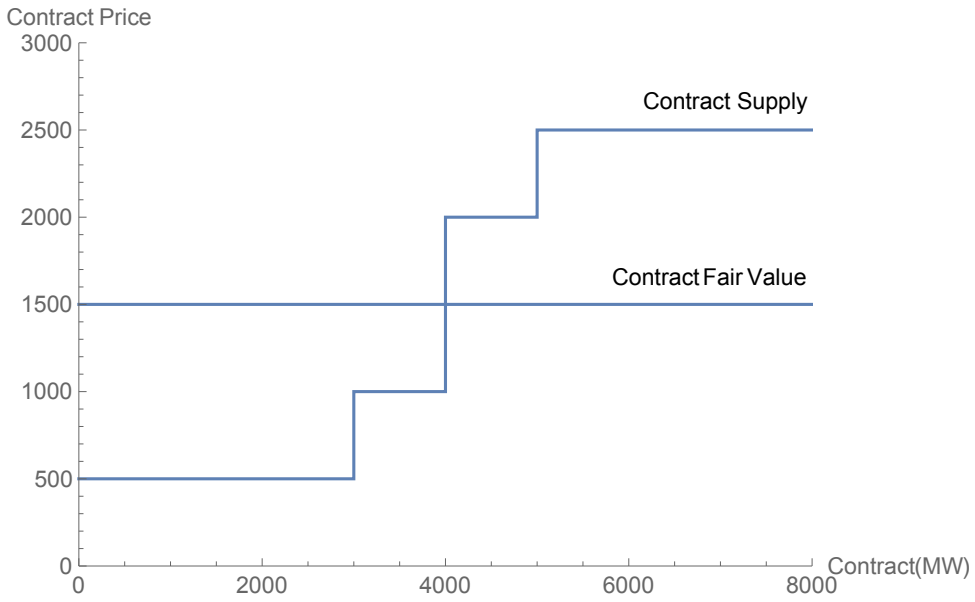



**Figure 39: Contract Supply**

The generator will be willing to sell contracts at a discount while they are under-contracted and the weighted average of spot market prices in the CVaR set is lower than $\lambda^{TWAP}$, the expected value of the contract. Conversely, generators will require a premium to fair value if they are to be enticed to sign contracts while already over-contracted, as additional contract signings increase, and not decrease, risk. The supply curve itself is also parameterised by the degree of risk aversion, and ranges from horizontal at fair value when the firm is not risk averse, to staggered and sensitive to CVaR scenarios. The premium or discount to $\lambda^{TWAP}$ increases with the degree of their risk aversion, as shown in Figure 40.

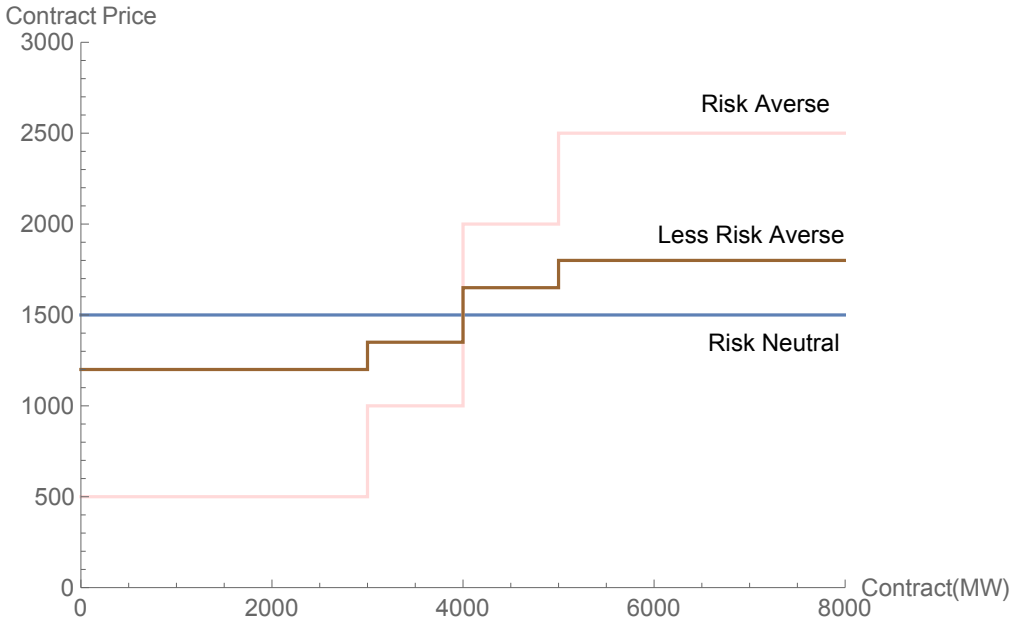



**Figure 40: Risk Aversion and Contract Supply**

*Consumers*

As with the previous discussion, we restrict our analysis to the choice to the same single contract, and ignore the potential for forming a portfolio of contracts to reflect a particular load profile. We assume the consumer's basic objective is to minimise the total purchase cost of a fixed energy consumption, which is defined by the sum of contract and spot purchase costs:

$$\sum_{s}\sum_{t} \mathrm{w}_{s,t} \sum_{r<R} \lambda_{r+1,s,t}\left(u_{r+1,s,t}-u_{r,s,t}\right)\left(\mathrm{L}-FWD^{d}\right)+\lambda^{C}FWD^{d} \tag{5.66}$$

The problem could be made more complex by consideration of the Load Weighted Average Price (LWAP) for each demand side participant, but this is just a different weighted cost and does not affect the analysis, other than creating a numerical difference, and providing potential justification for demand-side participants having different risk profiles. The complementarity conditions describing the consumer's optimal contracting policy are:

$$\lambda^{C}-\sum_{s}\sum_{t} \mathrm{w}_{s,t} \sum_{r<R} \lambda_{r+1,s,t}\left(u_{r+1,s,t}-u_{r,s,t}\right)\geq 0 \quad \perp \quad FWD^{d}\geq 0 \qquad \forall i \tag{5.67}$$

Recognising the summation expression is $\lambda^{TWAP}$ we have:

$$\lambda^{C}-\lambda^{TWAP}\geq 0 \quad \perp \quad FWD^{d}\geq 0 \qquad \forall i \tag{5.68}$$

In the absence of risk aversion, where the contract price exceeds $\lambda^{TWAP}$, the time weighted average price, a consumer will not demand any forward contracts. When the contract price is less than $\lambda^{TWAP}$, full contracting will occur. We now consider the case where the consumer, being risk averse, also applies a CVaR penalty, defined as:

$$CVaR^d = \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \left[ \lambda_s^{TWAP} \left( \mathrm{L} - FWD^d \right) + \lambda^C FWD^d \right] \tag{5.69}$$

Here $\alpha_s^d$ is determined in conjunction with $\alpha^{VaR,d}$, using the approach discussed in Section 5.4.1. The consumer's objective is:

$$\theta^d \sum_s \mathrm{w}_s \pi_s^d + \left(1 - \theta^d\right) \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \pi_s^d \tag{5.70}$$

Where:

$$\pi_s^d = \lambda_s^{TWAP} \left( \mathrm{L} - FWD^d \right) + \lambda^C FWD^d \qquad \forall s \tag{5.71}$$

The consumer's optimal contracting condition is governed by:

$$\theta^d \left( \lambda^C - \lambda^{TWAP} \right) + \left(1 - \theta^d\right) \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \left( \lambda^C - \lambda_s^{TWAP} \right) \geq 0 \quad \perp \quad FWD^d \geq 0 \tag{5.72}$$

Which can be more conveniently stated as:

$$\lambda^C - \left[ \theta^d \lambda^{TWAP} + \left(1 - \theta^d\right) \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \lambda_s^{TWAP} \right] \geq 0 \quad \perp \quad FWD^d \geq 0 \tag{5.73}$$

Risk-averse consumers will purchase forward contracts until:

$$\lambda^C = \theta^d \lambda^{TWAP} + \left(1 - \theta^d\right) \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \lambda_s^{TWAP} \tag{5.74}$$

This represents a convex combination of the time-weighted average over all scenarios and those over the worst scenarios as adjudged by the consumers CVaR optimisation. We can state the acceptable premium as:

$$\left(1 - \theta^d\right) \left[ \sum_s \frac{\alpha_s^d}{\alpha^{VaR,d}} \lambda_s^{TWAP} - \lambda^{TWAP} \right] \tag{5.75}$$

The consumer will be willing to pay a risk premium while they are under-contracted and the time-weighted average price in the CVaR scenarios is higher than $\lambda^{TWAP}$, the expected value of the contract. Conversely consumers will require a discount if they are to be enticed to sign contracts while already over-contracted, as additional contract signings increase, and not decrease, risk. The premium or discount to $\lambda^{TWAP}$ naturally increases with the degree of their risk aversion.

*Contract Market Clearance*

The clearance of the contract market is governed by the following complementarity condition:

$$\sum_g FWD^g - \sum_d FWD^d \geq 0 \quad \perp \quad \lambda^C \geq 0 \qquad \forall i > 0 \tag{5.76}$$

The simplest case to consider is that in which all participants are risk neutral. In this case, as shown in Figure 40, generator supply curves for contracts are perfectly elastic at the price $\lambda^{TWAP}$. From Figure 40, the same result holds for demand-side participants. The equilibrium is not unique in terms of quantity, but it is in terms of price, as while any contract quantity could be traded, trades will only occur at $\lambda^{TWAP}$. At that price, all participants are also indifferent as to whether they trade contracts or not. Where individual participants are not all risk neutral, the situation becomes significantly more complex. Aggregate demand and supply in the contract market are shaped by many factors, including the actual risks being faced, the degree of risk aversion, and the completeness or otherwise of the risk market. We begin by considering risk aversion while maintaining the completeness assumption.

Where participants are risk averse the demand function for contracts is decreasing, albeit not monotonically, and in the same fashion, the supply of contracts is increasing. As we assume market completeness, and the only contract available addresses price risk, we implicitly assume that price risk is the only risk in the market. Under those conditions, alteration of the contract quantity will only adjust the CVaR set once, at the balanced level where the contract and physical portfolio swap dominance from one to the other. Each individual demand and supply function has only two tranches. One corresponds to the contract levels below the contact quantity that would balance the portfolio, while the other applies to contract levels above the contract quantity that would balance the portfolio.

The degree to which each tranche deviates from $\lambda^{TWAP}$ is determined by the level of risk aversion, as shown in Figure 40. If we aggregate the demand for, or supply of, contracts, we get a piecewise constant function, for which the number of tranches is double the number of unique risk aversion levels among the participants involved. Following the supply-side, we know that for each generator, g, the contract price that corresponds to the level of risk aversion they experience is:

$$\lambda^{C} = \theta^{g}\lambda^{TWAP} + \left(1-\theta^{g}\right)\sum_{s}\frac{\alpha_{s}^{g}}{\alpha^{VaR,g}}\lambda_{s}^{TWAP} \tag{5.77}$$

At the contract quantity that balances their portfolio, the CVaR changes sign, and the acceptable contract price reflects a premium to fair value rather than a discount. The supply curve for contracts is vertical at this quantity. The same is true for the demand-side. The level at which supply and demand are balanced in aggregate is the same as expected total supply and demand in the underlying energy market are also equal in equilibrium. The vertical sections corresponding to a contract level equal to the expected quantity cleared in the energy market is defined by the contract price ranges in which the participant would achieve a balanced portfolio:

$$\theta^{g}\lambda^{TWAP} - \left(1-\theta^{g}\right)\sum_{s}\frac{\alpha_{s}^{g}}{\alpha^{VaR,g}}\lambda_{s}^{TWAP} \le \lambda^{C} \le \theta^{g}\lambda^{TWAP} + \left(1-\theta^{g}\right)\sum_{s}\frac{\alpha_{s}^{g}}{\alpha^{VaR,g}}\lambda_{s}^{TWAP} \tag{5.78}$$

Outside of this price ranges the demand or supply for contracts would be higher or lower. In the case of generators, at prices above the maximum of the range they will willingly supply an unlimited number of contracts, whereas below the minimum of the range, they will supply zero. To achieve equilibrium, the market-clearing price must satisfy (5.78) for all participants, d and g. In a complete

market, that implies the price must fall within the range defined by the least risk averse or, more generally, most risk neutral, participant. This example is in agreement with the more general result in Smeers & Ralph (2011). Accordingly, the risk premium/discount finally paid is set by the most risk-neutral participant, which is suggestive of the benefits of having properly functioning financial risk markets. We can further confirm this by re-arranging (5.78), for any pairwise comparison of participants:

$$\lambda^C = \lambda^{TWAP} + \frac{RP^S + RP^D}{2} \tag{5.79}$$

Where $RP^S$ and $RP^D$ are the equilibrium risk premiums for the generator and consumer respectively. From the perspective of the generator and the consumer we also have:

$$\lambda^C = \lambda^{TWAP} + RP^S = \lambda^{TWAP} + RP^D \tag{5.80}$$

The implication of (5.80) is to confirm that all participants pay equal post-contracting risk premiums in equilibrium, as we would expect with a unique price determination, and by virtue of contracting to the point where that price reflects the marginal benefit of contracting, they face the same marginal risk as the most risk neutral participant (Ralph & Smeers, 2011). The result is considerably narrower than that of Ralph & Smeers (2011) in that in this example all participants have precisely the same risk set, and there is only a single risk.

*Third Parties*

As forwards are financial contracts it is theoretically, and practically, possible for any institution capable of satisfying the market's prudential requirements to enter. Participants that are large and have significant diversification can operate with close to zero correlation so that in the limit they are almost entirely diversified. Where third parties are fully diversified, arbitrage will eliminate any differential between the contract price and the average spot market price. If a participant is risk neutral and without budget constraint, then the risk premium in equilibrium would be zero, as the risk neutral participants demand or supply for contracts is perfectly elastic at fair value and they would enter as many contracts as were made available by other participants at a suitable price. In general, risk markets do not necessarily deliver complete contracting, but the determination of risk pricing by the most risk neutral player shows the importance of having a risk neutral participant, such as a third party, who can trade risk premiums as close to zero as possible.

### 5.6.4 Contract Market Incompleteness

Contract markets are markets for trading risk that cannot be more cheaply controlled with internal risk management policies. Effective contract markets, are by definition, unlikely to develop where the distribution of stochastic variability is difficult to accurately quantify. In our framework those forms of variability are better treated as uncertainty, and not risk. However, even when the risks are well defined there is no guarantee that the risk markets available to investors are complete, and offer the opportunity to hedge all risk types and forms. A similar perspective is adopted in Boucher & Smeers (2012), in which the incompleteness of contract markets deprives the market of efficient outcomes.

Forward markets are too narrow in scope and duration to provide firms with the ability to hedge all risks. Additional contractual forms that address different needs and timeframes would advance the collective of risk markets towards completeness, but inevitably the transaction costs associated with contract formation and valuation necessitates that contract structures will be a compromise of a finite number of different needs, and risk markets will be destined to be incomplete to one degree or another in the face of many different risks. Market incompleteness can result from a number of issues.

*Contract Duration*

A very significant issue is the divergence between the time structure of risk and the time structure of contracts. Standardised contracts such as forwards leave investors exposed to risks of duration longer than the contract. Where contracts are signed for a duration that is smaller than that of the risk being contemplated, contracts will be continually re-valued against the backdrop of new information. Because a contract signed in the face of a particular risk will only provide a hedge for the duration of the contract, investors will remain uncontracted and concerned about long-term variability or permanent changes that might impact their profitability over the economic life of their investment.

*Uncontracted Income streams*

Alternative income streams are available in energy markets, such as for the provision of ancillary services for example. Contracts to manage the risk of this income stream may or may not be available. The effect on the firm is that a portion of profit variability remains uncovered by contracts. Where there is a bias between technologies, for example, in terms of their income generating activity, investment will be skewed away from those technologies that recover a higher proportion of their costs from alternative sources.

*Risk Market Distortions*

Just as in the spot market there are a number of strategic issues that can arise. For example, we have not considered in detail the possibility that generators may attempt to actively manipulate the risk measure of consumers, as suggested in Batstone (2000) in order to drive the risk premium paid by consumers higher. Instead, as with the spot market, we assume that in the long term the economics of entry will incentivise entry into either or both the contract market wherever risk adjusted returns justify investment. To assume otherwise would necessarily require discussion and rationalisation of a set of entry barriers that would protect such rent-seeking behaviour from competitive forces. While that is a worthwhile avenue of investigation and the subject of further ongoing research, it is beyond the scope of this work.

Given that markets are incomplete, the consequence for the analysis above is that, aside from coincidental outcomes, the most risk neutral participant will no longer necessarily define the price of risk. The contracting price range for each participant is defined as before but there is no longer a direct translation between the level of risk aversion and the price range as defined. For example, quantity risks are different for each participant and so the contract level at which each participant is balanced is different from their expected equilibrium spot market position. The market clearance, and associated

risk premium, are now defined by the participant with the lowest risk premium as defined by their supply or demand curve, which is dependent upon their underlying level of risk aversion and their unhedged position with respect to risks that are not addressed in the contract market, in this case non-price risk. Ralph & Smeers (2011) makes clear that the reason for this is not the consideration of the second type of risk per se, it is that the second type of risk is not traded. Were it traded, there would emerge a system risk agent who would assume the most risk neutral role, and set the price of risk in both markets for all participants.

## *5.7 Uncertainty*

We began by discussing capital recovery without risk aversion. In this case, capital recovery is based on earnings that match the amortised cost of capacity, which incorporates the interest costs of the firm, at a rate presumably assessed by a bank or financier. However, those interest costs reflect the banks assessment of their own risk, and not the risk faced by the firm. The introduction of risk aversion requires the firm to take account of the risks they take, and by valuing that risk, adjust decision making to achieve an optimal balance between risk and return, so that the extra profits earned in expectation cover the risks being taken. We progress a step further and consider uncertainty, so that the recovery of capital consists of its capital cost, a premium for the risks taken, and a premium for the uncertainty assumed by the firm.

Uncertainty may be unstructured and apply indiscriminately to the entire industry, or it may be structured in some way. We approach uncertainty from the perspective that all forms of variability can be at least partially explained by a probability distribution and that this information is valuable because ultimately we implicitly assume a distribution and/or determine the relevance of everything that is included, or not included, in our model (Farrar, 1964). So, uncertainty could refer to the residual doubt an investor has about a parameter estimate or a distribution, but it also encompasses the possibility of model misspecification through to the consideration of "black swan" events, which by definition are not predictable, and not modelled.

### 5.7.1 Formulating Uncertainty

One approach to addressing uncertainty is expanding the scenario tree to include modelling of the distribution of distributions, or even forming alternative scenario trees that might represent entirely different hypotheses about the underling structure of the market. But it is not simply a question of considering potential empirical inaccuracies in the modelled distribution of variability. Expansion of the formal modelling structure would do little to elucidate the position in terms that are readily understood or meaningful to investment equilibrium decisions as they are made in practice.

> *"The economist as such does not advocate criteria of optimality. He may invent them .....the ultimate choice is made by the procedures of decision making inherent in the institutions, laws and customs of society"*

In the long-term we would anticipate that the attitude of firms might equilibrate but the application of rational expectations with uncertainty may be rather fraught, particularly in the short-term, and requires further research.

To be logically consistent, equilibrium modelling must either assume uncertainty does not exist, and therefore no adjustment is required, or if it does exist, compensate for it. The former seems implausible, and the latter is the subject of little research, although uncertainty may have significant implications for the equilibrium plant mix, and the flow on metrics of pricing and generation adequacy by which the system judges itself. In fact, addressing uncertainty in an authoritative way is impossible as by definition it is that portion of the variability beyond that which we cannot explain with distributions. Nevertheless, investors and firms operate in an uncertain world and deal with uncertainty on a continuous basis, and the implication is that they do so on an ad hoc basis. Accordingly, we can either ignore the issue, or consider an analysis which is not geared towards advising a better strategy or containment mechanism, but which will improve understanding of the implications of uncertainty for market equilibrium, based on whichever conjectures that describe investor attitudes to uncertainty. We do not investigate the parameters or settings that may be applicable in a specific market, or even how to assess these, but we do consider some logical implications that the market structure might have on the form of the uncertainty adjustments made by firms.

When analysing risk we were able to attach penalties to the scenarios corresponding to the worst outcomes and as a result were able to construct a risk adjusted probability measure that amounted to skewing higher weightings to those scenarios corresponding to worse outcomes at the expense of de-weighting the scenarios corresponding to more favourable outcomes. Uncertainty provides no such basis for that form of adjustment, as it is not clear which scenarios will be affected more or less than others, and there is no way to be prescriptive with such unknowns. We conjecture that investors require some form of premium over and above the certainty equivalent that accounts for risk to account for unknown variability that defies specification, but in their experience has influenced outcomes in the past. We could think of this as a contingency premium.

The equilibrium investment condition can be re-arranged as follows:

$$\sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \le 1 \quad \perp \quad CAP_i \ge 0 \qquad \forall i > 0 \tag{5.81}$$

Investment based on this condition will recover enough additional profit in expectation to cover the imputed cost of the risk adjustment. Risk-averse investors will equate the certainty equivalent, and not the expected value, with the fixed costs of investing. The difference between the certainty equivalent and the expected return is referred to as a risk premium and, ceteris paribus, the premium is greater the more risk averse the investor is. We now consider a manager who desires a premium of r as a contingency in addition to the bare cost and risk premium recovery afforded by (5.81).

In itself this orientation is different to risk. Risk is assessed endogenously, with the actual level of risk that eventuates being an endogenous property of the equilibrium based on the variability

modelled and the risk aversion of the investor. In contrast, uncertainty is defined in terms of outcomes, requiring a certain premium as part of the definition of the equilibrium. Accordingly, with an uncertainty premium r, expressed as an additional rate of return requirement, the new equilibrium investment condition is:

$$\sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \leq 1+r \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{5.82}$$

Manipulation of (5.82) shows the burden of the risk premium is linear in fixed costs and therefore higher for plants with higher fixed costs. In standard complementarity form we have:

$$1+r-\sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{5.83}$$

Being a simple rate of return requirement, it might be thought that it can be incorporated like a risk adjustment to the discount rate used in determining the amortised equivalent of fixed costs. Unfortunately, as explained in Robichek & Myers (1966), the use of risk-adjusted discount rates would conflate the influence of uncertainty with adjustments in the time value of money. Another alternative is to express this as an adjustment to the risk neutral probabilities:

$$1-\sum_s \omega_s^{UNC} \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{5.84}$$

Where:

$$\omega_s^{UNC} = \frac{\theta}{1+r}\mathrm{w_s} + \frac{(1-\theta)}{1+r}\frac{\alpha_s}{\alpha^{VaR}} \qquad \forall s \tag{5.85}$$

Whereas the risk-neutral probabilities developed in Section 5.5.2 were a probability measure, when adjusting for uncertainty, the new weightings no longer sum to unity. Unlike risk, the effect of uncertainty is not to reapportion the emphasis placed on individual scenarios, but to degrade the influence of all scenarios. Accordingly, uncertainty produces a non-additive probability measure (Dow & Ribeiro da Costa Werlang, 2003). As shown in (5.85), de-weighting the scenarios associated with the CVaR measure might seem counter-intuitive as these are a measure of system risk, but uncertainty also de-weights other outcomes so that overall, given average profits exceed CVaR profits the value of the firms objective, or the investors return, is reduced by uncertainty. That reduction takes the form of scaling by $1/(1+r)$, and will flow through to investment and contracting decisions.

### 5.7.2 Utilisation Factors and Optimal Plant Mix

The impact of uncertainty of this form on the optimal plant, and on optimal trade-offs is clear. As the marginal profitability of a technology falls, capacity must be reduced to increase returns until parity between fixed costs and the uncertainty, and risk adjusted, returns is achieved. As shown in Figure 41, the optimal total generation capacity is lower.

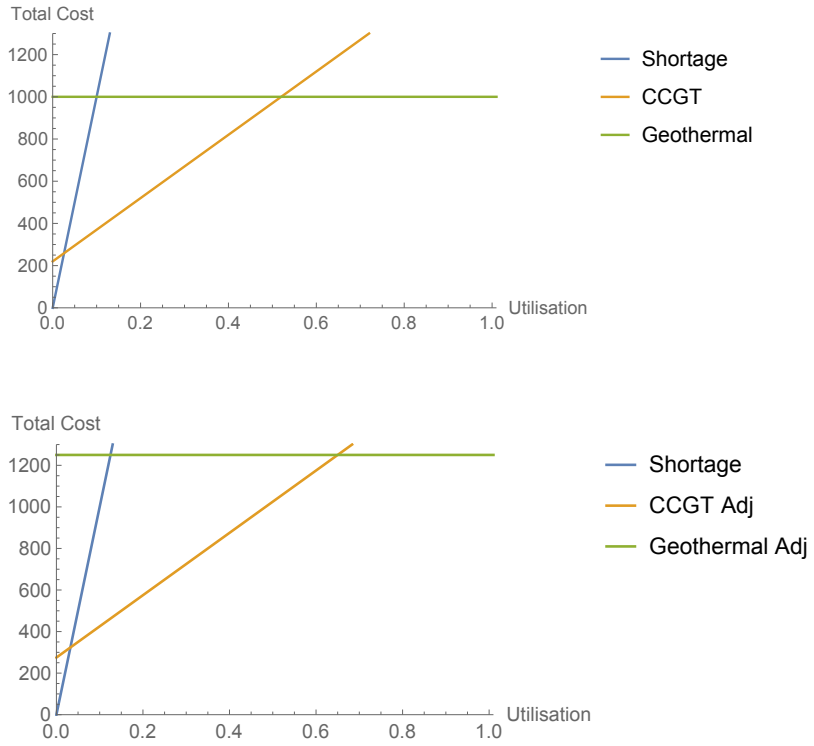



**Figure 41: Investment & Standard Business Uncertainty**

In relative terms, the technology mix is skewed away from high fixed cost technologies towards low fixed cost technologies, including the notional shortage technology. We can verify that by comparing the standard optimal trade-off definition and the optimal trade-off definition with an uncertainty adjustment:

$$u_{i,j} = \frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} < (1+r)\frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} = u_{i,j}^{adj} \qquad \forall i,j \qquad (5.86)$$

### 5.7.3 Structured Uncertainty

We have developed a formulation that incorporates and highlights two distinct approaches for dealing with risk and uncertainty. We have assumed uncertainty requires a premium to justify investment, and that this premium is assessed on the profits of the firm, affecting all parts of the firm equally. This can be interpreted as a sectoral uncertainty premium. We now assume that the firm adopts a more sophisticated and structured view of the required premium. On an ex-ante basis, placing a structure on uncertainty might be thought contradictory to the definition of uncertainty. But investors may have a relative lack of confidence in one or other aspect of the modelled results that leads them to discount certain aspects more than others.

*Uncertainty & Technology Type*

The unstructured approach to uncertainty suggests it is pervasive across the entire market, and the premium required can be summarised by a single constant parameter. But the investor may hold certain beliefs pertaining to specific technologies and apply a different premium to each. For example,

the required rate of return on nuclear technologies may be higher given the potential for accidents and the regulatory response to them that might cause interruption of their usage in the future.

$$1+r(i)-\sum_{s}\omega_{s}\frac{\pi_{i,s}}{\mathrm{FC_i}}\geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i>0 \tag{5.87}$$

Naturally, the effect of this form of uncertainty is to bias the plant mix away from technologies investors think have a higher degree of uncertainty attached to them.

*Uncertainty & Utilisation Factors*

The approach of the previous section potentially conflates the uncertainty associated with each technology and the uncertainty associated with its role. Following from Read (2005), entry economics dictate the equilibrium technology choice and the equilibrium utilisation level for that technology are connected, and it is therefore natural to confuse the equilibrium risk associated with a particular technology and the risk associated with its role in the plant mix. The same misunderstanding could apply with uncertainty. For example, an investor who might have experienced an unforeseen event that reduced peaking plant profitability could develop an uncertainty premium and attach that to the peaking technology, and not the peaking role. The two may well be the same in equilibrium, but penalising technologies for that which is not inherent in the technology, can potentially result in a technology with an assessed lower exposure to uncertainty being enlisted to perform the role of a technology with a supposedly higher exposure when that perception was rooted in the role they each typically performed.

$$1+r(u)-\sum_{s}\omega_{s}\frac{\pi_{i,s}}{\mathrm{FC_i}}\geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i>0 \tag{5.88}$$

By way of example, we conjecture that investors assess the exposure to uncertainty of a peaking role as being higher than that of a base load role.

This may be based on observance of the pattern of optimal cost recovery, which requires significant proportions of revenue be generated at relatively uncommon times of shortage, leaving technologies with lower utilisation factors proportionally more exposed to system shocks. This uncertainty may be based in doubts as to whether a regulator will allow windfall profits in the event of a crisis, even when those windfall profits are actually required to fund infrequently operated plants.

Boucher & Smeers (2012) develops an example where this might be the case, in the context of security of supply issues in Europe. Without formally relating the issue to the strict interpretation of uncertainty, they suggest that contracts for low probability and high cost events will not come to fruition. They also note that even when arrangements do evolve to fund these sorts of costs, investors may harbour doubt about the ability of the government to confront politicisation of the issue without succumbing to market intervention. In either scenario, the investor might well be justified in treating the situation as "uncertain", and accordingly discounting the associated revenue further.

We adopt the following function to reflect a hypothetical uncertainty adjustment across the utilisation range:

$$r(u) = r(u_0)m^{1-u} \tag{5.89}$$

This functional form has some desirable properties for illustrative purposes, but many other forms, including a linear or piecewise linear form could be used. We choose m to define the uncertainty premium of the role of ultimate peaker relative to the base-load role, whose uncertainty premium is given by $r(u_0)$. For, example if m=2, we would apply double the uncertainty premium to operations with a very low utilisation factor than would be applied to base load generation roles. We can define the optimal trade-off between technologies as follows:

$$u_{i,j} = \frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} < \left(1 + r(u)\right)\frac{\mathrm{FC_j} - \mathrm{FC_i}}{\mathrm{MC_i} - \mathrm{MC_j}} = u_{i,j}^{adj} \qquad \forall i, j \tag{5.90}$$

Uncertainty of this form brings an increase in shortage frequency as the breakeven points between each technology all increase when uncertainty premiums are applied. Further, the relative size of the movement implied by each type of uncertainty adjustment at each utilisation level is different as the uncertainty adjustment is decreasing in utilisation.

The application of a sector wide uncertainty adjustment has little impact on technologies with low capital costs so, for example, the change in shortage frequency is relatively small because OCGT capital costs are small. At the opposite end of the plant spectrum, that form of adjustment results in a large absolute increase in annualised capital costs and large reductions in the optimal capacity of the plant involved. Penalising by utilisation factor does not discriminate between technologies, so while the shortage frequency and breakeven points move right under both approaches, a utilisation based adjustment subjects those technologies in peaking roles to a far greater risk adjustment than those technologies filling a base load role. Accordingly, the implication is that we observe relatively greater increases in shortage frequency, and larger decreases in utilisation of peaking plant than if we simply applied a sector wide uncertainty factor.

In separating technological and utilisation-based uncertainty, we have not denied existence of the former. There are likely to be differences between technologies that extend beyond those that are associated with the equilibrium utilisation of the technology. A general form of expression could be defined as:

$$1 + r_i(u) - \sum_s \omega_s \frac{\pi_{i,s}}{\mathrm{FC_i}} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \tag{5.91}$$

Where:

$$r_i(u) = r_i(u_0)m_i^{1-u} \tag{5.92}$$

This formulation allows individual base uncertainty adjustments and progressions for each technology.

The framework presented in the rest of this thesis would adapt to uncertainty expressed in this fashion by simply replacing the equilibrium investment constraint with whichever variant of uncertainty adjustment is desired. The adjustment of optimal trade-offs is expressed in terms of a single period model, and identifies those utilisation levels where the trade-offs between various

technologies exist. However, the definition of optimal trade-offs in the multi-period framework is based on sub-period profitability and so the definition of optimal trade-offs in our model need not be altered.

The inclusion of uncertainty in a conventional optimisation with fixed exogenous utilisation levels will only further cloud the determination of spot market pricing. In those models, spot market prices would be directly disciplined by uncertainty as this wold be accounted for in cost recovery price setting. The approach taken in this thesis limits the influence of uncertainty, and risk for that matter, to the influence that is transmitted through capacity choices. That a uncertainty surrounding a sunk investment can influence spot market prices is extremely inconsistent.

## *5.8 Summary and Conclusions*

In this chapter, we have implemented a CVaR risk measure to consider risk. The conditional probability distribution that defines CVaR is not based on the dual of a standard optimisation, and is found by direct maximisation of the worst case amongst the risk set of the investor. The KKT conditions of that optimisation form part of the model. Accordingly, unlike many implementations of CVaR that focus on a broader range of the distribution and measure all but the most favourable outcomes, our definition of CVaR is direct and addresses the risky set of the investor directly. Whereas the former approach involves the optimisation of CVaR alone, our approach is designed to be combined with expected outcomes. The combination of expected returns and a risk measure avoids the need to maintain an irrational preference structure in which investors are assumed to be indifferent to all scenarios not included in the calculation of CVaR.

Having formed the objective function as a convex combination of expected profits and CVaR, in Section 5.5.3 we generalise this approach to include multiple CVaR definitions. This enables sculpting of the return distribution, and the application of different weightings to areas of the distribution that might have different real world implications for the investor. This approach requires the construction of a separate optimisation, and therefore a separate set of complementarity conditions, for each CVaR measure. Although not necessarily the case, the interpretation of CVaR in this case lends itself to consideration of multiple agents, servicing different aspects of the risk profile faced by the investor.

Section 5.5.4 examines the equilibration of the system utilising different examples. The combined expected profit and risk measure can be viewed as the convex combination of two PDC's, one with risk neutral probabilities, the other corresponding to the risk set that defines CVaR. We observe that as capacity is added each PDC adjusts, and as the PDC defines the marginal benefit of investment, so does the slope of that function. We also note that when the risk set changes, there will be a discrete change in the corresponding PDC, and the slope of the marginal benefit will also change.

We introduce contracting in Section 5.6. While significantly simplified, our approach is aligned with that taken in Ralph & Smeers (2011,2015). Ours is a complete market with endogenous pricing of risk, in which participants play a risk averse Nash game. We derive the form of the demand and supply curve and find the most risk neutral participant is price setting, and by extension all other

participants observe the same price of risk. This is in alignment with the result in the paper above. We finish our discussion of contracting by considering incomplete markets, in which the above result no longer holds, and the reason for markets being incomplete.

Finally we discuss uncertainty, as a distinct concept from risk. In the literature, these terms are used interchangeably. Our objective is illustrate a distinction between the terms that is based on whether the distribution for some stochastic variability is known or unknown. We argue that there is a significant difference between investor responses in each case. In the case where the distribution of future events is known, this is risk, and based on the knowledge of that distribution it is likely that risk markets can form and operate and that risk can be hedged to some degree or other. When the distribution of future scenarios, or even the scenarios themselves are not known, we have uncertainty. Our approach to modelling the impact of uncertainty focuses on the investment equilibrium condition.

We begin with a fixed uncertainty premium, representative of an investor believing that they need an additional premium above the hedged return distribution they believe they will receive, to account for anything unconsidered, or misunderstood in the model. The implications are higher returns are required and a bias against high fixed cost technologies in the plant mix. We then consider a more nuanced approach, in which the investor perceives uncertainty does not exist uniformly across the system. First we consider the case where uncertainty is greatest at higher utilisation levels, perhaps because of the potential for unspecified strategic intervention in times of high price. In this case, the adjustment we propose an adjustment function that makes the risk adjustment a function of the utilisation of a particular technology. The subtlety in this adjustment lies in the fact that often risk penalties are assessed on technologies based on the typical role they play, whereas in this case we assess the same penalty on all technologies in a given role, and let equilibration determine what the ultimate return profile of the technology is. Nevertheless, there is also good reason for assessing uncertainty on the basis of technology, and we do this, allowing a different function for each technology.

Throughout this chapter, the implications of integration with our framework are discussed. The integration of risk aversion requires wholesale adjustment to the equilibrium investment condition. The method of determining optimal trade-offs remains unaffected as these are based on imputed values, although the values themselves will naturally change. The calculation of at least one, and potentially more, CVaR measures requires the addition of a set of complementarity conditions for each. Contract market clearance also requires a further set of constraints, and depending on the nature of demand-side modelling, this could generate another set of complementarity constraints corresponding to the optimisation problem faced by consumers. Finally, the introduction of uncertainty merely involves adjustment of the investment equilibrium constraint.

The definition of CVaR is direct and we are unaware of this form in any investment model. The model itself uses a convex combination of expected profits and risk and is a contribution itself, following an avenue of further research suggested in Ehrenmann & Smeers (2011). As far as we are aware, the consideration of multiple CVaR measures in a single investment model is also novel. Our contribution in terms of contracting is limited to confirming for our highly simplified example that the

general result in Ralph & Smeers (2011) holds, and that contracting can be integrated into our framework.

Insofar as uncertainty is concerned, much of the literature has focussed on dealing with uncertainty through the use of sensitivity analysis, but this does not address the unforeseen, or unforeseeable, and perhaps more importantly, it is not an equilibrium concept. Our contribution is very preliminary, but we have shown how uncertainty adjustments could be integrated into the general framework, so that endogenous uncertainty, or strategic uncertainty could rightfully be considered as next steps in this investigation.

# 6 SUMMARY AND CONCLUSIONS

At its commencement, the goal of this research was to investigate a number of interactions in the electricity market with a view to understanding investment and the equilibration of the market. This process began by considering the type of framework that would be suitable for such an enterprise and primarily for the benefit of future research, it was decided that complementarity models would be the vehicle for this. As the framework developed, we became concerned at the discrepancy between conventional optimisation formulations, and screening curve analysis, in which the latter seemed more accurate, and the former more adaptable. As a result, the resolution of these issues and the development of the framework consumed the available space in this thesis. Nevertheless, the thesis has made contributions in a number of areas and provides a basis for further research in the areas initially identified as its focus.

The research contained herein can typically be viewed as relating to one of the following over-arching themes through present in this thesis. Broadly speaking they are:

- The economic relationships and equilibration processes in electricity markets, understood at a fundamental level.
- The use of complementarity as an analytical framework, capable of uniting many mathematical methods.

The analysis began with a review of investment fundamentals and, in particular, optimisation models and screening curve analysis. Screening curve analysis, while conceptually strong is computationally limited. As a result, practitioners migrated to mathematical programming to analyse the investment and generation problem. That the conventional optimisation formulations they developed are not consistent with the logic of screening curves is problematic and requires investigation, and resolution.

The thesis provides that resolution, but not before coming to terms with the root cause of the inconsistency. An initial concern was the staging of the model. Ostensibly, conventional optimisation formulations treat investment and generation activities as being simultaneous. That issue proved not to be central to the problem observed. In Chapter 1 we demonstrate this by presenting a two stage formulation of the problem, which can be readily restated as a single stage formulation, on account of the second stage being a sub-gradient of the first stage. The single stage version of those formulations matches the problem at hand, whereas we show by example that conventional optimisation formulations related to the general formulation do not in a number of ways. The distinguishing feature of these formulations is that they are specialisations of the general formulation, that rely on a definition of the LDC and specification of the functional form of generation functions. We provide formal models with piecewise constant, piecewise linear, and higher order LDC forms, each of which have advantages and disadvantages in their own right. The common feature they all have though, is that in a conventional optimisation, the break points of the generation functions in those models are restricted to the breakpoints used in the definition of the LDC. This can be thought of as an arbitrary restriction on generation functions, or the optimal utilisation levels of each technology.

The consequence of this anomaly is that even when the LDC specification is absolutely precise, the solution of the conventional formulation is sub-optimal and not a competitive equilibrium. We show this by example, and through demonstration of its Pareto-inefficiency. We also show that the PDC defined is inconsistent with the capacity prescribed by the model solution. Of practical importance, the extent of the inaccuracy in the primal capacity decision can be significantly greater than the error in the objective function, which may provide a false sense of security. The mathematical requirement of cost recovery is achieved in the conventional optimisation by including uplift components in pricing. In general, these are not available through the spot market clearing process. Aside from the impact of arbitrary restrictions on generation functions and utilisation levels, we noted a particular case where the economic structure itself, implicitly imposes the same restrictions. This is the case when the LDC is piecewise constant, and results in ambiguous pricing over a subset of load classes, with no particular price level having precedence over another.

The resolution of the problem comes from integrating the logic of screening curves with optimisation. Broadly speaking we add the utilisation corresponding to optimal technological trade-offs to the LDC definition. This task becomes endogenous in more complex models so our approach must address both the identification of optimal trade-offs, and the integration of these amongst other utilisation levels included for the purpose of LDC definition. We rely on option pricing principles to define endogenous capacity values to ensure utilisation and the PDC are consistently represented when global technological trade-offs do not apply, as in sub-periods for example. Use of endogenous dual values throughout the thesis allows a consistent specification of optimal trade-offs, even while the trade-offs themselves may be changing.

In Chapters 3 we demonstrated adaptations of our framework designed to consider issues that may be of relevance in electricity markets today. From the perspective of investment analysis, the emphasis in each chapter was to capture the flavour of particular extensions at a level which an investor might be interested. We considered technological generalisations such as non-constant cost structures, capacity limits, and energy limits with a few to demonstrating the integration of relatively standard features in our framework. The introduction of configurable technologies was a significantly more complex exercise, and lead to the development of discrete marginal operating ranges for each technology, generalising the traditional single optimal trade-off in screening curve analysis.

In Chapter 4 we investigated issues that involve the endogenous determination of the LDC. Our approach sought to define short and long term demand response separately. This distinction is important, and as short term demand response is capable of being price setting, while long term demand response is not, we remove the possibility that the latter can be marginal in the spot market. Our examination of reliability followed the conventional approach by augmenting load by expected outages and assuming perfect reliability thereafter. To complete the system, returns had to be scaled by the reliability of the technology concerned. Unsurprisingly, unreliability leads to additional capacity and cost, although the implications for individual technologies depend on their relative reliability. In both of the preceding cases, the modelling was simplified by the monotonic relationship between load and price, and load and outages, respectively. This simplification is not available when considering intermittent generation. To accurately capture the earnings of intermittent technologies we must

consider the correlation of intermittent generation with net load. On the basis that intermittent generation may have a daily pattern, we specified that correlation chronologically, on the basis of daily chronological load. Total intermittent generation is endogenous and determined by intermittent capacity levels. Net chronological load, accounting for the possibility of spillage, is then available by deduction for the chronological load profile. Finally the net LDC is created by re-ordering the chronological profile, and applying a scale factor that represents the additional variability in the LDC that is not present in chronological load, which is a point estimate. The net LDC is serviced by conventional technologies that define the PDC.

Finally, we turn to risk, which is a significant issue in investment. While CVaR optimisations are available, we elected to define CVaR directly by forming the dual of the traditional formulation, and treating conditional probabilities as variables. This enables the combination of the CVaR risk measure with expected profits, and we take advantage of that possibility by promoting the possibility of sculpting the return distribution with multiple CVaR measures assessed at different significance levels, or with narrower focus in the scenario tree. When considering contracting, our approach was limited to showing that this could be done in the framework. Nevertheless, there are a number of ways to include contracting, but ultimately we followed the path of Ralph & Smeers (2011), defining the risk market clearance endogenously. Stochastic endogenous equilibria provide the most desirable basis for future research. Insofar as our narrow implementation can, it reinforced one of the key results form that paper, that with complete risk markets the price of risk is defined by the most risk-neutral participant. We also introduce uncertainty in this chapter. The purpose is to create a distinction with risk. We consider several forms of uncertainty adjustment, clarifying the difference between technologies and their typical use. We reiterate that uncertainty adjusted scenario weightings do not sum to one, as uncertainty amounts to a penalty on the whole, rather than a reallocation between the parts.

The preceding discussion primarily relates to the first theme of the thesis: the development of fundamental economic features of electricity markets in our framework. In each case that we discussed in Chapters 3,4, and 5, we identified the necessary changes in our framework to accommodate the feature under discussion. The flexible definition of optimal trade-offs established earlier meant that the form of this complementarity condition changed little throughout the thesis. As far as the basic investment problem was concerned, more attention was required to ensure that the investment condition was updated. In each of those cases, conventional optimisation formulations could not solve the problems presented for the same reason that they fail to solve the simplest case as in each case, and each sub-period or scenario, the optimal solution requires correct specification of optimal trade-offs and a correct assessment of the PDC, both of which elude conventional optimisation formulations as we have shown. While the determination of utilisation levels and the representation of the PDC are essential from the perspective of the basic framework and are what enables a problem to be solved where it could not with conventional optimisation formulations, the most significant adjustment required in each case was the addition of considerable numbers of complementarity conditions to provide the infrastructure necessary to represent the issue of concern.

At this point, we address the second major theme of the thesis: the use of complementarity models as an analytical framework. In economic problems, conditions in complementarity problems

typically arise from the KKT conditions of optimisations. This can be seen in our framework as KKT conditions from investment optimisation and spot market clearing problems define the basis for this. As we discovered, complementarity models are significantly more flexible than this. Perhaps the first distinction worthy of note is the inclusion of optimisations that do not relate to a market participant. Examples include the ranking optimisation that combines endogenous and exogenous utilisation levels, the scale factor determination for the LDC in the section on intermittent generation, or the definition of CVaR, although the latter is readily identifiable as some form of actual agent if desired. The KKT conditions corresponding to these optimisation sit as equals amongst those that arise from actual participants and agents. These problems can also be nested with complementarity conditions and this was done to achieve the minimal set of critical utilisation levels, in which each step is an optimisation defined whose starting point is defined by the solution to the previous optimisation.

Complementarity conditions also can be expressed directly, and represent logic. The definition of optimal utilisation levels is an example of a set of conditions that defines and bounds an endogenous variable. These bounds can also be used creatively to filter solutions. In the case of configurable technologies, the difference of two quadratics is naturally quadratic, although it may have no real roots, implying no intersection. By carefully constructing proxies for the various components of the quadratic equation, we were able to modify complex solutions into solutions that were real, and with the help of a scaling factor, large, so that they would be ignored by the rest of the formulation. For the ultimate purpose of integrating intermittent generation, a similar approach was used to effectively toggle the denominator and numerator for the expression defining utilisation within a chronological segment, so that we avoided division by zero and the definition was suitable when load was increasing and decreasing. In combination, the thesis demonstrates the power of complementarity theory to represent multiple optimisation objectives, nested problems, logical constructs and, in combination, algorithmic approaches to identifying solutions.

The specific contributions are generally noted throughout the thesis but the following highlights some specific contributions that, to our knowledge, are original:

- Endogenous definition of utilisation levels in an investment model
- Development of (multiple) methods for pruning optimal trade-offs
- Endogenous definition of capacity values in individual scenarios and sub-periods
- Modelling an endogenously determined range of continuously configured technologies, including the determination of multiple pairwise technological trade-offs
- Endogenous modelling of intermittent generation using both chronological and duration load forms, with dynamic construction the net LDC based on investment decisions.
- Modelling of a weighted expected value and CVaR objective function in electricity market investment model (inspired by Ehrenmann & Smeers 2011) including the use of a direct optimisation of CVaR.
- Sculpting of the return distribution with multiple CVaR constraints in an electricity investment model

Opportunities for further research abound. There are significant technological changes causing re-assessment of investment decision making in electricity markets. Firms, regulators and governments each face strategic decisions beyond those typically modelled, yet potentially as important. Furthermore, relative to risk, the impact of uncertainty is not well understood, and the source of uncertainty, and the endogenous creation or dissipation of uncertainty should be considered. Lastly, aside from advancing directly to the consideration of the sort of interactions envisaged at the beginning, there is the over-riding question of determining the most appropriate means to solve these sorts of formulations.

# 7 APPENDICES

## *7.1 Example Implementation of the Conventional Approach*

### 7.1.1 Problem Description

The objective of the problem is to minimise the total costs of servicing load. The total costs include capital as well as variable costs.

### 7.1.2 Problem Solution

The solutions are compared below:

| Technology | Conventional Approach | Thesis Approach |
|---|---|---|
| Notional Shortage | 0 | 696 |
| OCGT | 12000 | 11637 |
| CCGT | 0 | 4333 |
| Coal | 20000 | 10333 |
| Geothermal | 58000 | 63000 |

The total cost of capacity and generation using the conventional approach is 78.35M. Using the thesis approach that total cost is lower at 77.43M.

### 7.1.3 Solution Methodology

The conventional optimisation formulation was solved using the included optimisation tools supplied with Microsoft Excel.

The actual optimal solution was found using the methodology of Chapter 2. In 2.3.6, we define the necessary sets of conditions to state this approach more generally, in a form that can be expanded as additional complexity arises. Re-capping, the approach requires:

- Definition of optimal trade-offs (2.4)
- Definition of critical optimal trade-offs (2.5). These are the ones that are important in practice, that apply between adjacent technologies.
- Integrated ordering of both the optimal trade-offs and those points necessary for LDC definition (2.6.1)
- Determination of load levels for utilisation levels corresponding to optimal trade-offs (2.6.2)

From Section 2.4.1, the optimal utilisation levels are automatically defined. These are then processed further to create a minimal set of optimal trade-offs:

| Technology | Optimal Trade-Off |
|---|---|
| Notional Shortage/OCGT | 0.0043 |

| | |
|---|---|
| OCGT/CCGT | 0.1133 |
| CCGT/Coal | 0.2867 |
| Coal/Geothermal | 0.7000 |

These can then be ordered amongst the utilisation levels included for the purpose of defining the LDC. Together with the load interpolations that correspond to the utilisation levels above, and in addition to those points that define the LDC, we have the definition of an identical LDC, with additional utilisation levels.

| Load (MW) | 90000 | 89304 | 82000 | 78000 | 77667 | 73333 | 63000 | 58000 | 50000 |
|---|---|---|---|---|---|---|---|---|---|
| Utilisation | 0.000 | 0.0043 | 0.050 | 0.100 | 0.1133 | 0.2867 | 0.7000 | 0.900 | 1.000 |

The remainder of the problem is a complementarity formulation that in this simple case is equivalent to the conventional optimisation formulation, with the addition of the necessary utilisation levels so that the generation functions, while restricted, are restricted to a set of breakpoints that includes the optimal breakpoints. This optimised version of the problem was solved as a linear program using Microsoft Excel, as shown below.

**Conventional Optimisation Formulation**

| Cost Structure | | | | LDC Definition | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Technology | Fixed Cost | Var Cost | | Load | 90000 | 82000 | 78000 | 78000 | 58000 | 50000 | | | | | | | | | | |
| Notional Shortage | 0 | 15000 | | Utilisation | 0 | 0.05 | 0.1 | 0.1 | 0.9 | 1 | | | | | | | | | | |
| OCGT | 50 | 3500 | | | | | | | | | | Energy Content by Load Class | | | | | | | | |
| CCGT | 220 | 2000 | | Load Classes | 1 | 2 | 3 | 4 | 5 | | | 1 | 2 | 3 | 4 | 5 | | | | |
| Coal | 650 | 500 | | Constant | 82000 | 78000 | 78000 | 58000 | 50000 | | | 4300 | 4000 | 0 | 54400 | 5400 | | | | |
| Geothermal | 1000 | 0 | | Duration | 0.05 | 0.05 | 0 | 0.8 | 0.1 | | | | | | | | | | | |
| | | | | Gradient | 160000 | 80000 | | 25000 | 80000 | | | | | | | | | | | |
| | | | | | Generation at Each Point | | | | | | | Energy by Load Class | | | | | | | | |
| Technology | Capacity | Cost | | Technology | 1 | 2 | 3 | 4 | 5 | 6 | | 1 | 2 | 3 | 4 | 5 | Total | Cost | | |
| Notional Shortage | 0 | 0 | | Notional Shortage | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| OCGT | 12000 | 600000 | | OCGT | 12000 | 4000 | 0 | 0 | 0 | 0 | | 400 | 100 | 0 | 0 | 0 | 500 | 1750000 | | |
| CCGT | 0 | 0 | | CCGT | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Coal | 20000 | 13000000 | | Coal | 20000 | 20000 | 20000 | 20000 | 0 | 0 | | 1000 | 1000 | 0 | 8000 | 0 | 10000 | 5000000 | | |
| Geothermal | 58000 | 58000000 | | Geothermal | 58000 | 58000 | 58000 | 58000 | 58000 | 50000 | | 2900 | 2900 | 0 | 46400 | 5400 | 57600 | 0 | | |
| | | 71600000 | | Total | 90000 | 82000 | 78000 | 78000 | 58000 | 50000 | | 4300 | 4000 | 0 | 54400 | 5400 | 68100 | 6750000 | | **78350000** |

**Thesis Approach**

| Successive Trade-Offs | | LDC | Capacity | LDC Definiton | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Notional Shortage | 0.0043 | 89304.35 | 695.65 | Load | 90000 | 89304 | 82000 | 78000 | 78000 | 77667 | 73333 | 63000 | 58000 | 50000 | | | | | | |
| OCGT | 0.1133 | 77666.67 | 11637.68 | Utilisation | 0 | 0.0043 | 0.05 | 0.1 | 0.1 | 0.1133 | 0.2867 | 0.7000 | 0.9 | 1 | | | | | | |
| CCGT | 0.2867 | 73333.33 | 4333.33 | | | | | | | | | | | | | | | | | |
| Coal | 0.7000 | 63000.00 | 10333.33 | Load Classes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | | | | | | |
| Geothermal | 1.0000 | 50000.00 | 63000.00 | Constant | 89304 | 82000 | 78000 | 78000 | 77667 | 73333 | 63000 | 58000 | 50000 | | | | | | | |
| | | | | Duration | 0.0043 | 0.0457 | 0.0500 | 0.0000 | 0.0133 | 0.1733 | 0.4133 | 0.2000 | 0.1000 | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | Generation | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | | | | |
| | | | | Notional Shortage | 696 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | |
| | | | | OCGT | 11638 | 11638 | 4333 | 333 | 333 | 0 | 0 | 0 | 0 | 0 | | | | | | |
| | | | | CCGT | 4333 | 4333 | 4333 | 4333 | 4333 | 4333 | 0 | 0 | 0 | 0 | | | | | | |
| | | | | Coal | 10333 | 10333 | 10333 | 10333 | 10333 | 10333 | 10333 | 0 | 0 | 0 | | | | | | |
| | | | | Geothermal | 63000 | 63000 | 63000 | 63000 | 63000 | 63000 | 63000 | 63000 | 58000 | 50000 | | | | | | |
| | | | | | 90000 | 89304 | 82000 | 78000 | 78000 | 77667 | 73333 | 63000 | 58000 | 50000 | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | Energy Content | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total | Cost | | Capacity Cost | | | |
| | | | | Notional Shortage | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 22684 | | 0 | | | |
| | | | | OCGT | 51 | 365 | 117 | 0 | 2 | 0 | 0 | 0 | 0 | 534 | 1869151 | | 581884.058 | | | |
| | | | | CCGT | 19 | 198 | 217 | 0 | 58 | 376 | 0 | 0 | 0 | 867 | 1733333 | | 953333.333 | | | |
| | | | | Coal | 45 | 472 | 517 | 0 | 138 | 1791 | 2136 | 0 | 0 | 5098 | 2548889 | | 6716666.67 | | | |
| | | | | Geothermal | 274 | 2876 | 3150 | 0 | 840 | 10920 | 26040 | 12100 | 5400 | 61600 | 0 | | 63000000 | | | |
| | | | | Total | 391 | 3912 | 4003 | 4 | 1043 | 13093 | 28183 | 12108 | 5409 | 68100 | 6174058 | | 71251884.1 | | | **77425942** |

## *7.2 Solution Ambiguity*

The nested optimisations in the algorithm from Section 2.5.2 determine the screening curve envelope. For each technology i in sub-period t, the utilisation level at which that technology becomes active in that algorithm is given by:

$$u_{i,t}^{\max} = \sum_{n=0}^{N-1} z_{i,n,t} u_{n,t}^{e} \qquad \forall i,t \qquad (7.1)$$

As the algorithm progresses from low utilisation levels to high utilisation levels, this utilisation level corresponds to the utilisation level below which the technology becomes inframarginal, as opposed to marginal. Were there no exogenous utilisation levels we could generate a restriction using this utilisation level, in the knowledge that the next utilisation level encountered would also be endogenously chosen and therefore the one at which technology i first enters the dispatch. Unfortunately, there are exogenous utilisation levels to be considered, so it is not the case that the next utilisation level corresponds to the point at which the technology becomes marginal in terms of dispatch. To define that point we modify the expression in (7.1) to define the next endogenous utilisation level:

$$u_{i,t}^{*} = \sum_{n=1}^{N} z_{i,n-1,t} u_{n,t}^{e} \qquad \forall i,t \qquad (7.2)$$

To resolve the ambiguity problem in a fashion consistent with screening curve analysis, we require $GEN_{i,r,t} = 0$ at all $u_{r,t} \geq u_{n,t}^{*}$. For the technologies involved in each optimal trade-off, satisfaction of this restriction requires the technology with the lowest marginal cost must supply capacity and generate the incremental load, $L_{r,t} - L_{r+1,t}$. It remains to introduce complementarity conditions to give effect to this restriction:

$$a_{i,r,t}^{0} + \left( \sum_{n=1}^{N} z_{i,n-1,t} u_{n,t}^{e} - u_{r+1,t} \right) \geq 0 \quad \perp \quad a_{i,r,t}^{1} \geq 0 \qquad \forall i,r,t \qquad (7.3)$$

$$a_{i,r,t}^{1} + GEN_{i,r,t} \geq 0 \quad \perp \quad a_{i,r,t}^{0} \geq 0 \qquad \forall i,r,t \qquad (7.4)$$

If $u_{r,t} \geq u_{n,t}^{*}$, then $u_{r+1,t} \geq u_{n,t}^{*}$. The latter holds as a strict inequality, except when $u_{r+1,t} = u_{r,t}$, in which case the incremental load $L_{r,t} - L_{r+1,t} = 0$, and from the standard market clearing constraint we have $GEN_{i,r,t} = 0$. Where $u_{r+1,t} > u_{n,t}^{*}$, the bracketed term in (7.3) is negative, implying $a_{i,r,t}^{0} > 0$ to ensure feasibility of (7.3). From (7.4), and with $a_{i,r,t}^{1} \geq 0$ and $GEN_{i,r,t} \geq 0$ it must be the case that we must have $GEN_{i,r,t} = a_{i,r,t}^{1} = 0$. So when $u_{r,t} \geq u_{n,t}^{*}$ we have $GEN_{i,r,t} = 0$ as desired. Taking the counter-case, we have $u_{r,t} < u_{n,t}^{*}$. In this case the bracketed term in (7.3) is non-negative as either $u_{r+1,t} < u_{n,t}^{*}$ or $u_{r+1,t} = u_{n,t}^{*}$. Where the bracketed term is non-negative there are enough degrees of freedom to allow $a_{i,r,t}^{0} \geq 0$ and by extension $a_{i,r,t}^{1} \geq 0$ and $GEN_{i,r,t} \geq 0$, although not all combinations of these restrictions can be achieved simultaneously. Where $u_{r,t} < u_{n,t}^{*}$, the market clearing conditions optimally require generation by technology I, so that generation by technology i is feasible, unlike the case where

$u_{r,t} \geq u^*_{n,t}$ in which generation by technology i is restricted to zero to enforce the logic of the screening curve and thereby remove all ambiguity.

## *7.3 Stochastic Energy Limits*

The problem of stochastic energy limits is essentially one of stochastic reservoir management, which is applicable to many fields of research, not the least of which is the wider energy and resource field. Within the research of electricity systems, reservoir management has received considerable focus on account of the significance of hydroelectric generation in many systems. The optimal management of reservoirs has been addressed with linear programming, but primarily with dynamic programming in a variety of forms (Read & Hindsberger, 2010), (Lino, Barroso, Pereira, Kelman, & Fampa, 2003; Pereira & Campodónico, 1999). The defining features of stochastic reservoir management that align it with dynamic programming are the stochastic nature of inflows, and the storage or reservoir which provides the opportunity to redistribute inflows across time. By their very nature, and our choice of terminology, stochastic inflows are generally associated with hydro systems, although future energy storage options may make this discussion appropriate for other natural energy sources. Ours is a necessarily modest foray into this area, based on the complementarity form developed here. The purpose of such a model in an investment equilibrium framework is to capture the difference between deterministic and stochastic decision making in a market with significant reliance on this generation, not to develop a detailed model of reservoir management.

We begin by introducing the stochastic inflow pattern. In each of the t-sub-periods there is a distribution of H(t) possible inflows quantities, $\mathrm{INF}_{t,h(t)}$, occurring with probability $p^{inf}_{t,h(t)}$. The storage at the beginning of period t is represented by a distribution of L(t) potential storage levels, $STOR_{t,l(t)}$, each one of which is dependent on previous release decisions. These occur with probability $p^{stor}_{t,l(t)}$, which is dependent on the stochastic inflow pattern up to that sub-period. We have not considered inflow correlation as contemplated by Yang (1999) in a dynamic programming context, but this could be included by defining $p^{inf}_{t,h(t),l(t)}$ so that the probability of each inflow h(t) is a function of past inflows. In each sub-period the index of possible beginning of period storage is defined as $l(t) = 1.....L(t)$, where $L(t) = \prod_{j=1}^{t} H(j)$. It is clear from the geometric progression of the size of the state space that this approach is not suitable for granular modelling when either the number of sub-periods is high, or the number of possible inflow options in each sub-period is high. A Markov decision process is a viable modelling alternative for larger problems provided that the state space, which is at least two dimensional, can be efficiently discretised without also falling prey to the curse of dimensionality. For small problems or conceptual analysis, this basic approach, which ignores the specific equilibration of opening and closing storage distributions, has the advantage of not requiring discretisation of the space, while being able to deliver a relatively rich inflow and storage structure, should this feature be an important aspect of a particular market.

We define the relationship between beginning of period storage and end of period storage by adapting (3.72) to account for multiple inflows and end of (previous) period storage levels:

$$STOR_{i,t-1,l(t-1)} + \mathrm{INF}_{\mathrm{i,t,h(t)}} - REL_{i,t,l(t-1)} - STOR_{i,t,h(t),l(t-1)} = 0 \quad \perp \quad \gamma_{i,t,h(t),l(t-1)} \; free$$

$$\forall i > 0, t > 0, l(t-1) = 1...L(t-1), h(t) = 1...H(t) \tag{7.5}$$

For each storage level $STOR_{i,t-1,l(t-1)}$, of which there are L(t-1) for each technology at the end of the previous sub-period, a single release decision, $REL_{i,t,l(t-1)}$, must be chosen before inflows are realised. There are H(t) possible inflow realisations, creating H(t)L(t-1) possible end of period storage levels for a given release decision. This creates expansion of the dimensionality of storage so we take the opportunity to flatten this dimension to provide clarity:

$$STOR_{i,t,h(t),l(t-1)} - STOR_{i,t,L(t\text{-}1)(l(t\text{-}1)\text{-}1)+h(t)} = 0 \quad \perp \quad \gamma_{i,t,l(t)} \; free$$

$$\forall i > 0, t > 0, l(t) = l(t-1) \times h(t) \tag{7.6}$$

In turn this implies:

$$\gamma_{i,t,h(t),l(t-1)} - \gamma_{i,t,l(t)} = 0 \quad \perp \quad STOR_{i,t,l(t)} \; free \quad \forall i > 0, t, l(t) \tag{7.7}$$

The indexing is flattened from a two-dimensional index to a single dimension index l(t)=1…L(t) using the expression in (7.6). From (7.5) and (7.6), $\gamma_{i,t,h(t),l(t-1)} = \gamma_{i,t,l(t)}$ and continues to represent the marginal value of stored fuel at the end of sub-period t.

Having defined the basic inflow structure, the principles guiding the management of storage and fuel releases are unchanged by stochasticity. The owner of the resource still seeks to maximise the value of the resource, while negotiating storage limits and stochastic inflow patterns. In the deterministic case the perfect allocation would equalise the marginal value of storage across all sub-periods, although that is not always possible when storage is finite. In the stochastic case the marginal value of stored fuel is determined recursively. Beginning with an end condition that we discuss later, the marginal value of stored fuel is recursively defined by marginal stored fuel values working backwards from the end of the period. The operator has a choice between releasing fuel in the current period or storing it for future use, and achieving the expected marginal value of stored fuel:

$$\sum_{h(t)=1}^{H(t)} \mathrm{p}^{\mathrm{inf}}_{\mathrm{i,t,h(t)}} \gamma_{i,t,h(t),l(t-1)} - \varepsilon_{i,t,l(t-1)} \geq 0 \quad \perp \quad REL_{i,t,l(t-1)} \geq 0 \quad \forall i > 0, t, l(t-1) \tag{7.8}$$

The expectation and release terms are parameterised by the system state l(t) and expectations are taken over the possible inflow sequences, h(t), to arrive at the expected marginal value of stored fuel as understood by the system operator under nonanticipativity. In equilibrium, where the expected marginal value of stored fuel exceeds the marginal value of releasing fuel, the release of fuel for generation in the current period is zero. Conversely when the marginal value of release exceeds the expected marginal value of stored fuel, there is an incentive to increase releases until parity is achieved.

Storage at the end of each sub-period must also be feasible. Where inflows disappoint, minimum storage levels must still be respected, and similarly where inflows are high, maximum storage limits must be respected. Accordingly, the storage bounds (4.70) and (4.71) must apply to all possible combinations of opening storage levels and inflow eventualities, as indexed by l(t):

$$STOR_{i,t,l(t)} - \mathrm{STOR}^{-}_{\mathrm{i,t}} \geq 0 \quad \perp \gamma^{-}_{i,t,l(t)} \geq 0 \qquad \forall i > 0, t \tag{7.9}$$

$$\mathrm{STOR}^{+}_{\mathrm{i,t}} - STOR_{i,t,l(t)} \geq 0 \quad \perp \gamma^{+}_{i,t,l(t)} \geq 0 \qquad \forall i > 0, t \tag{7.10}$$

We can describe the inter-temporal linkages between optimal marginal stored fuel values with the following condition:

$$\gamma_{i,t,l(t)} - \sum_{h(t)} \mathrm{p}^{\mathrm{inf}}_{\mathrm{i,t+1,h(t+1)}} \gamma_{i,t+1,h(t+1),l(t)} + \gamma^{+}_{i,t,l(t)} - \gamma^{-}_{i,t,l(t)} = 0 \quad \perp STOR_{i,t,l(t)}\ free \quad \forall i > 0, t, l(t) \tag{7.11}$$

The operator decides on a fuel release strategy at the beginning of each sub-period, before inflows are known. Inflows occur with a given probability and, along with the release decision, generate a probability distribution for the value of end of sub-period storage, which carries forward to the next sub-period. The goal of the risk neutral operators is to equate the beginning of period marginal value of stored fuel with the expected end of period marginal value of stored fuel as much as is possible while respecting storage limits. Where storage limits are not binding, then for each end of sub-period storage level, $STOR_{i,t,l(t)}$, the marginal value of stored fuel at the end of sub-period t is equal to the expected marginal value of stored fuel at the end of sub-period t+1.

Where the operator is risk averse, the nature of the risk adjustment is dependent on the nature of their risk aversion, or put another way, the specific constraints that might be placed on the operational decisions of the reservoir manager. It is possible that the reservoir manager wishes to consider the distribution of returns alluded to in (7.11), rather than just the average return. In this case, a risk management structure would need to be introduced to resolve the manager's view of a given distribution of returns. While possible, we consider it more likely that the distribution of annual profits is a more natural concern of the firm. From one perspective, the risk-neutral losses will be augmented by penalties and therefore higher. Accordingly, the future value of the resource will be elevated in those scenarios, and this incentivises increased retention of fuel for future usage. From another perspective, the effective weighting attached to undesirable scenarios is increased, affording the value of fuel in those scenarios a greater role, and encouraging reservoir management policies to place greater emphasis on managing those scenarios. In each case, the effect of the risk management policy is introduced implicitly to (7.11), in one approach through adjustments to future fuel values, while in the other the adjustment is through changing probabilities.

In the final sub-period, we have a set of outcomes with an associated probability distribution. We can recursively express the cumulative probability of being in each state l(t) at time t as follows:

$$\mathrm{p}^{\mathrm{stor}}_{\mathrm{i,t,l(t)=L(t-1)(l(t-1)-1)+h(t-1)}} = \mathrm{p}^{\mathrm{stor}}_{\mathrm{i,t,l(t-1)}} \mathrm{p}^{\mathrm{inf}}_{\mathrm{i,t,h(t-1)}} \qquad \forall i > 0, t, l(t) \tag{7.12}$$

The expected marginal value of stored fuel at the end of the period is:

$$\sum_{l(T)} \mathrm{p}^{\mathrm{stor}}_{\mathrm{i,T,l(T)}} \gamma_{i,T,L(T)} \qquad \forall i > 0 \qquad (7.13)$$

As in the deterministic case, we need to ensure that the process of fuel consumption is sustainable, and we accomplish this by using the function $V_i^{/}\left(STOR_{i,t,l(t)}\right)$ as the marginal value function:

$$\gamma_{i,t,l(t)} - \sum_{h(t)} \mathrm{p}^{\mathrm{inf}}_{\mathrm{i,t+1,h(t+1)}} \gamma_{i,t+1,h(t+1),l(t)} \Big|_{t<T} - \sum_{h(t)} \mathrm{p}^{\mathrm{inf}}_{\mathrm{i,t+1,h(t+1)}} V_i^{/}\left(STOR_{i,t,l(t)}\right) \Big|_{t=T} + \gamma^{+}_{i,t,l(t)} - \gamma^{-}_{i,t,l(t)} = 0$$

$$\perp \; STOR_{i,t,l(t)} \; free \qquad \forall i > 0, t, l(t) \qquad (7.14)$$

As in the deterministic case, the progression of the marginal value of stored fuel through sub-periods is based on maintaining equivalence between beginning of period marginal values and the expected end of period marginal values as in (7.11). Overall discipline is provided by (7.14) which ties the expected marginal value of stored fuel to the marginal value of releases in each sub-period. Extreme inflow sequences will be characterised by marginal storage values that will have significantly diverged from average levels. Depending on the direction of the deviation, fuel will either be relatively abundant or scarce in these scenarios, and the influence that releases have on the marginal value of stored water in these conditions will be greater for that divergence, which will lead to release strategies that promote moderation.

The dimensionality of the problem is significantly expanded if we consider stochastic inflow patterns and releases are now defined not just for each sub-period but for each system state, based on a given beginning of period storage level of l(t), in sub-period t.

$$REL_{i,t,l(t-1)} - \frac{1}{2} \sum_{r=0}^{R-1} \left[ \left( GEN_{i,r,t,l(t-1)} - GEN_{i,r+1,t,l(t-1)} \right) \left( u_{r+1,t,l(t-1)} + u_{r,t,l(t-1)} \right) \right] \geq 0$$

$$\perp \qquad \varepsilon_{i,t,l(t)} \geq 0 \qquad \forall i > 0, t, l(t) \qquad (7.15)$$

The market clearing is also significantly expanded to account for the dimensionality, so that there is a market clearance in each sub-period corresponding to each beginning of period system state, as indexed by l(t-1). Along with the equilibrium investment condition, the conditions determining optimal trade-offs remains the same, although each are also in higher dimensions.

## *7.4 Ancillary Services*

### 7.4.1 Introduction

Ancillary services are designed to provide system security within the dispatch timeframe. Section 4.3 deals with the slightly overlapping issue of reliability, but apart from the actual dispatch period in which an unscheduled breakdown or outage occurs, the reliability discussion relates to medium term reliability or "firmness", although we acknowledge that in certain markets, reserve can be called upon to prevent shortages. Our focus here is on how reserve provision specifically relates to the intra-dispatch timeframe. From an investment point of view, ancillary services have some important

implications. Ancillary service markets provide another source of income for those technologies able to provide services, introduce valuation differentials between different plant types based on their reserve providing capabilities, and complicate the decision and offer processes of firms who must choose between competing uses.

### 7.4.2 Formulating Reserve Provision

*Instantaneous Reserve*

Perhaps the most common ancillary service procured is instantaneous reserve. Instantaneous reserve is designed to protect the system and ensure continuance of supply when a unit, plant or transmission line fails. Reserve of this sort is typically available in several classes, based on the technical capabilities of each technology, such as their ability to respond, and sustain that response. Accordingly, in NZ for example, there is a six second and sixty second reserve class.

Depending on the system being studied, the level of instantaneous reserve required will be set to ensure at least the largest single outage is covered (N-1 rule), and in some markets the largest two outages are covered (N-2 rule). The reserve requirements for the market are often endogenous to the market clearance process itself, but can also be set on an exogenous basis:

$$\sum_i RES_{i,r,t} - RES^{up}_{r,t} \geq 0 \quad \perp \quad \lambda^{res}_{r,t} \geq 0 \qquad \forall r,t \tag{7.16}$$

Setting reserve requirements on an endogenous basis, involves taking account of specific plant output and transmission flows to determine the level of reserve required as a function of the optimal dispatch. In NZ, the level of reserve required in each island is determined by the largest of the maximum generation from a single plant in each major island, and the net reduction in transfer across the inter-island HVDC link island, due to pole failure. Others have proposed more elaborate methods for determining a dynamic reserve (Mitropoulos, 1984)

Not all technologies are equally adept at providing instantaneous reserve, and some may not be able to at all. The capacity of each technology available to provide instantaneous reserve is defined in (7.16), which demonstrates that capacity must ultimately be allocated to one purpose or another. In addition, there is a maximum amount of reserve that can be provided due to the ability of plant to ramp up output in the specified time-frame, which is represented here as a constant, but one which can be optimised at the time of plant design, or refurbishment. That relationship is unlikely to be convex, but we approximate the ability of a plant to provide reserve with the following constraints, using them to describe the convex approximation of the set of feasible operations.

$$\tau^{ramp}_i GEN_{i,r,t} - RES_{i,r,t} \geq 0 \quad \perp \quad \mu^{ramp}_{i,r,t} \geq 0 \qquad \forall i,r,t \tag{7.17}$$

$$\tau^{+}_i CAP_i - RES_{i,r,t} \geq 0 \quad \perp \quad \mu^{+}_{i,r,t} \geq 0 \qquad \forall i,r,t \tag{7.18}$$

$$CAP_i - GEN_{i,r,t} - RES_{i,r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,r,t} \geq 0 \qquad \forall i,r,t \tag{7.19}$$

If it were technically possible, reserve would be provided by plant that was not in operation, and this would not reduce capacity available for energy production. But as described by condition (7.17),

reserve is typically only available from plants already operating, with capacity for reserve supplied by backing off plant, starting at the top of the operating merit order and then progressing down the merit order until reserve constraints are satisfied. This becomes increasingly expensive, in opportunity cost terms, as lower marginal cost energy producing technologies earn higher energy market profits. The cost of providing reserve therefore depends on the market equilibrium, as reserve provision is priced against the opportunity cost of foregoing production opportunities.

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} - \tau_i^{ramp}\mu_{i,r,t}^{ramp} + \varphi_{i,r,t}^{+}\Big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \tag{7.20}$$

When the ramping constraint on reserve provision is inactive, (7.20) collapses to the standard form. However, when the ramping constraint is binding an increase in generation also affords an increase in reserve provision, effectively subsidising the marginal cost of generation. The complementarity condition corresponding to reserve provision is:

$$-\lambda_{r,t}^{res} + \mu_{i,r,t}^{ramp} + \mu_i^{+} + \varphi_{i,r,t}^{+}\Big|_{i>0} \geq 0 \quad \perp \quad RES_{i,r,t} \geq 0 \qquad \forall i,r,t \tag{7.21}$$

If the case where technology i is providing reserve, then from (7.17) we have $GEN_{i,r,t} > 0$ so that combining (7.20) and (7.21) gives:

$$-\lambda_{r,t}^{res} + \mu_{i,r,t}^{ramp} + \mu_i^{+} + \varphi_{i,r,t}^{+}\Big|_{i>0} = -\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} - \tau_i^{ramp}\mu_{i,r,t}^{ramp} + \varphi_{i,r,t}^{+}\Big|_{i>0} \tag{7.22}$$

Re-arranging gives the relationship between energy market and reserve pricing:

$$\lambda_{r,t}^{res} = \lambda_{r,t} - \left(\mathrm{MC}_{\mathrm{i,t}} - \tau_i^{ramp}\mu_{i,r,t}^{ramp}\right) + \mu_{i,r,t}^{ramp} + \mu_i^{+} \tag{7.23}$$

When neither the ramping nor the maximum reserve constraints are binding, the equilibrium reserve price reflects the opportunity cost of the devoting capacity to the energy market. When reserve constraints bind, the reserve price is higher than spot market profitability in equilibrium, as we would expect incentives remain to swing capacity to the reserve market while the firm is constrained from doing so. The ramping constraint dual reflects the cost of that constraint in the reserve market, but also the effective subsidy of marginal cost provided when generation is increased.

At certain utilisation levels reserve pricing could entice generators to partially exit the spot market and divert capacity to the reserve market. In a perfectly competitive market, the opportunity cost of providing reserve is the profitability of the technology, so the reserve price will be equal to the profitability of the individual technology supplying reserve at the margin in any case. As that technology must be producing and marginal (partially utilised) to provide reserve, the opportunity cost of that technology supplying reserve is lower than the profitability of a fully utilised technology entering lower in the merit order.

The difficulty with determining critical utilisation levels in this case is that the cost functions of each technology in the energy market are piecewise linear with the effective discount to accommodate the ramping constraint being represented as a reduction in that technologies marginal cost. Were the cost curve piecewise linear, this problem would be easily solved, but the added degree of difficulty in this case is a result of the definition of the piecewise linearity being endogenous and a

property of the equilibrium. In some cases it could be possible for the benefit available from supplying reserve to exceed the cost of operating at loss in the spot market and could lead to a technology being willing to offer below cost in order to be dispatched and collect reserve payments. That would be represented by a downward sloping total cost curve at utilisation levels higher than those for which the technology would be inframarginal.

We have also not considered the reserve implications and the implicit requirement for flexibility in other technologies, even over multiple dispatch periods, further complicating the assessment of the cost of intermittent generation. There is some concern that the price depressive effect of wind and other intermittent generation forms will impact most negatively on those technologies that would have the flexibility to assist with its integration (Traber & Kemfert, 2011), (A. Wu, 2012).

## *7.5 Demand Response as a Configurable Technology*

For each pairwise comparison of technologies there is a load response available between them that is based on the price differential between the two technologies. The quantum of load response available depends on the properties of the demand curve and the marginal cost differential between the technologies being compared. The total load response is set by the demand function, and system prices, and so jointly contributes to, and is a property of the equilibrium. The actual load response in the market is a function of which technologies ultimately define market clearances. We proceed by defining the capacity and functional form of pairwise load response opportunities so that these may be selected and define critical utilisation levels, capacities and prices, using an adaptation of the algorithm as described in Section 3.5, modified to consider load response.

Whereas in the case of considering configuration options the orientation used to define the quadratic portion of the cost structure was automatically established by the use of known limiting technologies, one of which was known to be more capital intensive than the other. It is not as clear with load response, because the value function of a load response opportunity is arrived at by comparing the cost structure of two different technologies, and so the direction of comparison is not known, and may not even be consistent between sub-periods or scenarios if, for example, energy limits are applicable. Fortunately, this is of no import, as the definition of the quadratic cost structure is invariant to the orientation of the trade-off specification.

Under perfect competition without load response, the spot price is determined by the marginal cost of the marginal generator. Load response occurs between those price levels which are defined by adjacent generation technologies. If we consider the lower marginal cost technology, then in order to preserve the option value of investment in this technology at the level of FC, the utilisation range occupied by demand response must be evenly drawn from the marginal operating range of this technology and the adjacent higher marginal cost technology. In order to retain this balance we impose constraint (7.24), in terms of $u_{n,t}^{lr,-}$ and $u_{n,t}^{lr,+}$, which describe the optimal utilisation range load response at this point. We note that this constraint is specific to a linear demand form.

$$u^{e}_{n,t} - \frac{u^{lr-}_{n,t} + u^{lr,+}_{n,t}}{2} = 0 \qquad \forall i \in LR, t \qquad (7.24)$$

As the piecewise linear LDC is monotone decreasing in utilisation, progressive extension of the utilisation range width $u^{lr-}_{n,t} - u^{lr,+}_{n,t}$, also monotonically increases the capacity corresponding to that utilisation range. Although the utilisation range is symmetrically extended, the same cannot be said for capacity, which will vary asymmetrically according to the slope of the LDC at the relevant utilisation level. Resorting to a simple linear demand function we have the following expression for the capacity of demand response available between technologies i and j, remembering that, in this instance, the load response operates around a critical utilisation level, $u^{e}_{n,t}$:

$$\begin{aligned} CAP^{lr}_{n,t} &= \left(\mathrm{L}^{0}_{\mathrm{t}} - \mathrm{a}_{\mathrm{t}}\lambda_{n-1,t}\right) - \left(\mathrm{L}^{0}_{\mathrm{t}} - \mathrm{a}_{\mathrm{t}}\lambda_{n,t}\right) \\ &= \mathrm{a}_{\mathrm{t}}\left(\sum_{j} z_{j,n}\mathrm{MC}_{\mathrm{j}} - \sum_{j} z_{j,n-1}\mathrm{MC}_{\mathrm{j}}\right) \end{aligned} \qquad \forall n,t \qquad (7.25)$$

Here $\mathrm{a}_{\mathrm{t}}$ is the price coefficient in the demand function, which when multiplied by the difference between adjacent marginal costs gives the capacity of load response, $CAP^{lr}_{n,t}$. This is the actual quantum of load response available between in this price range and not merely a bound. The utilisation range of the load response opportunity will expand until the capacity as defined by the LDC reaches the required level defined in (7.25).

Demand response technologies are parameterised by the marginal costs of adjacent technologies and the total quantum of load response available which is also a function of those marginal costs, along with the demand function, and the LDC. Based on a linear marginal value of load response as determined by the demand function, the value of load response is a quadratic expression. From (3.80), we have a definition of the optimised cost of generation with a configurable technology with linear fixed and marginal cost adjustment. We can adapt this to define a LRV (Load Response Value) function that is analogous to a total cost function for a generation technology:

$$LRV_{i,t}(u) = FC^{0}_{i,t} + \left(fc_{i,t} + MC^{0}_{i,t}\right)u - mc_{i,t}u^{2} \qquad \forall i \in LR, t \qquad (7.26)$$

We introduce $FC^{0}_{i}$ and $MC^{0}_{i}$ to describe both the intercept and the slope of the load response value function at zero utilisation. In this context, these are parameters and have lost a degree of interpretability as a result of no longer also describing the limiting version of a technology. $FC^{-}_{i,t}$ and $FC^{-}_{i,t}$ still define bounds on the value of load response in the range bounded by $MC^{-}_{i}$ and $MC^{+}_{i}$, the marginal costs of the two adjacent technologies. $fc_{i,t}$ and $mc_{i,t}$ continue to describe the rate of change the intercept and slope of the function in linear terms. Just as with configurable technologies, in which there is a continuous spectrum of trade-offs, in this case there is a continuous range of consumer load response, with each increment in the price encouraging a slight reduction in demand.

The quadratic coefficients are jointly determined by the fixed and marginal costs of adjacent technologies, $MC_i^-$ and $MC_i^+$, and utilisation levels which are disciplined by the total available capacity. We begin by expressing $fc_{i,t}$ and $mc_{i,t}$ in terms of other variables and known parameters:

$$\mathrm{mc}_{\mathrm{i,t}} = \frac{\mathrm{MC}_{\mathrm{i}}^{+} - \mathrm{MC}_{\mathrm{i}}^{-}}{u_{i,t}^{+} - u_{i,t}^{-}} = \frac{MC_i^0 - \mathrm{MC}_{\mathrm{i}}^{+}}{u_{i,t}^{-}} \qquad \forall i \in LR, t \qquad (7.27)$$

$$\mathrm{fc}_{\mathrm{i,t}} = \frac{\mathrm{FC}_{\mathrm{i}}^{+} - \mathrm{FC}_{\mathrm{i}}^{-}}{u_{i,t}^{+} - u_{i,t}^{-}} = \frac{\mathrm{FC}_{\mathrm{i}}^{-} - FC_{\mathrm{i}}^{0}}{u_{i,t}^{-}} \qquad \forall i \in LR, t \qquad (7.28)$$

We define the slope of (7.26) as follows:

$$\frac{\partial TC_i(u)}{\partial u} = \mathrm{fc}_{\mathrm{i}} + \mathrm{MC}_{\mathrm{i}}^{+} - 2\mathrm{mc}_{\mathrm{i}} u \qquad \forall i \in LR, t \qquad (7.29)$$

It follows that at $u_{i,t}^-$ we have:

$$\mathrm{fc}_{\mathrm{i,t}} + \mathrm{MC}_{\mathrm{i,t}}^{+} - 2\mathrm{mc}_{\mathrm{i,t}} u_{i,t}^{-} = \mathrm{MC}_{\mathrm{i,t}}^{+} \qquad \forall i \in LR, t \qquad (7.30)$$

Substituting suitably chosen definitions of $fc_{i,t}$ and $mc_{i,t}$:

$$\frac{\mathrm{FC}_{\mathrm{i}}^{-} - FC_{\mathrm{i}}^{0}}{u_{i,t}^{-}} + \mathrm{MC}_{\mathrm{i,t}}^{+} - 2\left(\frac{MC_i^0 - \mathrm{MC}_{\mathrm{i}}^{+}}{u_{i,t}^{-}}\right) u_{i,t}^{-} = \mathrm{MC}_{\mathrm{i,t}}^{+} \qquad \forall i \in LR, t \qquad (7.31)$$

Solving for $u_{i,t}^-$ yields:

$$u_{i,t}^{-} = \frac{\mathrm{FC}_{\mathrm{i}}^{-} - FC_{\mathrm{i}}^{0}}{2\left(MC_i^0 - \mathrm{MC}_{\mathrm{i}}^{+}\right)} \qquad \forall i \in LR, t \qquad (7.32)$$

A similar expression is available for $u_{i,t}^+$:

$$u_{i,t}^{+} = \frac{\mathrm{FC}_{\mathrm{i}}^{+} - FC_{\mathrm{i}}^{0}}{2\left(MC_i^0 - \mathrm{MC}_{\mathrm{i}}^{-}\right)} \qquad \forall i \in LR, t \qquad (7.33)$$

Re-arranging expressions (7.32) and (7.33) uniquely determines the parameters $FC_{\mathrm{i}}^0$ and $MC_i^0$, that complete the definition of $LRV_{i,t}(u)$. As with configurable technologies, this response is only locally valid at $u_{i,t}^- \le u \le u_{i,t}^+$. This information will form part of the criteria for selecting critical utilisation levels. In the case of generation technologies the adjustment rates of both the fixed and marginal cost are known, whereas in this case we know only the adjustment rate of the marginal value, and the overall capacity available. In conjunction with the LDC these define an endogenous intercept which describes the value of each incremental load response to the system, and which is consistent with the total quantum of load response available. In that sense load response is valued individually in each

sub-period or scenario, and its capacity, while not unlimited, will vary according to the prevailing prices and technological interactions.

### 7.5.1 Critical Utilisation Levels

Unfortunately the process of selecting critical utilisation levels is significantly more complex. Previously we stepped through the lower envelope starting with a notional shortage technology, and progressively selecting the minimal optimal trade-offs with the current technology, before moving to the implied new technology and repeating the process until it reached a conclusion.

Where load response is concerned, if we imagine the complementarity conditions in algorithmic form, we are unable to simply choose the minimal next optimal trade-off whether that it corresponds to a generation technology or a load response option. If the currently selected technology is a generation technology then by definition any transition in marginal cost will create a load response opportunity, so we must move from generation technology to load response opportunity. If the current technology is a load response opportunity from between technologies i and j, then this load response must intersect next with technology j before any other for it to even be valid.

To assist in enforcing the alternating pattern we introduce the following complementarity constraint:

$$ALT_n + ALT_{n-1}\big|_{n>0} - 1 = 0 \quad \perp \quad ALT_{n-1} \; free \qquad \forall n \tag{7.34}$$

The primary constraint of the original formulation defined the next critical utilisation level by selecting the minimum intersection from all such intersections with other technologies j:

$$u_n^e = \sum_j z_{j,n} \sum_i \mathrm{z}_{\mathrm{i,n\text{-}1}} \mathrm{u}^{\mathrm{e}}_{\mathrm{i,j}} \qquad : \psi_n^0 \tag{7.35}$$

$$\sum_j z_{j,n} = 1 \qquad : \psi_n^1 \tag{7.36}$$

As we commence with the notional shortage technology we wish to alternately select a technology from the set of load response options and then generation technologies. Accordingly we modify (7.35) so that the minimum value of $u_n^e$ is given by:

$$u_n^e = ALT_n \sum_j z_{j,n} \sum_i \mathrm{z}_{\mathrm{i,n\text{-}1}} \mathrm{u}^{\mathrm{e}}_{\mathrm{i,j}} + \left(1 - ALT_n\right) \sum_j z^{\mathrm{lr}}_{j,n} \sum_i \mathrm{z}^{\mathrm{lr}}_{i,n-1} \mathrm{u}^{\mathrm{lr}}_{\mathrm{i,j}} \qquad : \psi_n^0 \tag{7.37}$$

We can also condense these constraints so that they only apply when required

$$\sum_j z_{j,n} = 1 \qquad : \psi_n^1 \tag{7.38}$$

$$\sum_j z^{lr}_{j,n} = 1 \qquad : \psi_n^1 \tag{7.39}$$

To prevent cycling we previously ensured that the marginal cost of the next technology should be no higher than the last, and that they should not be repeated. This approach remains reasonable although the actual technologies are separated by load response opportunities. By ensuring that generation technologies are not repeated, the same is implied for load response opportunities.

$$\sum_j z_{j,n\text{-}1}\mathrm{MC}_j - \sum_j z_{j,n}\mathrm{MC}_j \geq 0 \quad \perp \quad \psi_n^2 \geq 0 \qquad \forall n > 0 \qquad (7.40)$$

$$-\sum_j z_{j,n\text{-}1} z_{j,n} + \psi_n^5 \geq 0 \quad \perp \quad \psi_n^3 \geq 0 \qquad \forall n > 0 \qquad (7.41)$$

$$\sum_j z_{j,n}{}^2 - 1 \geq 0 \quad \perp \quad \psi_n^4 \geq 0 \qquad \forall n \qquad (7.42)$$

$$1 - u_n^e \geq 0 \quad \perp \quad \psi_n^5 \geq 0 \qquad \forall n \qquad (7.43)$$

## *7.6 Stochastic Dominance*

Stochastic dominance refers to the screening of different investment options to find the most preferred, or dominant choice. As shown in Figure 42, the curve denoted FSD exhibits first order stochastic dominance over the base case, as the return is higher in every eventuality. The curve denoted SSD exhibits second order stochastic dominance over the base case. Although it fails to dominate in every eventuality, the expected returns up to any level are greater for this distribution than they are for the base distribution. Mathematically, first order stochastic dominance is defined as:

$$F_i(y) \geq F_j(y) \ \forall y \qquad (7.44)$$

Second order stochastic dominance is defined as:

$$\int_{-\infty}^{z} F_i(y)\,dy \geq \int_{-\infty}^{z} F_j(y)\,dy \ \ \forall z \in y \qquad (7.45)$$

In practice, FSD will not effectively screen many options, as real decisions usually involve some form of trade-off between two outcomes that are, at least in a pairwise sense, Pareto-efficient in terms of risk and return. If, in addition to accepting the reasonableness of FSD, we are also prepared to accept the firm is risk averse we can apply a screening based on second order stochastic dominance (SSD). If the firm exhibits decreasing risk aversion, third degree stochastic dominance will screen out possible options based on that criterion. Going beyond third order stochastic dominance, while possible, is difficult to credibly align with underlying economic assumptions.

Although it is considerably less common than other approaches, the analysis of risk and decision making with stochastic dominance has been operationalised in the literature. Bunn (1984) summarises stochastic dominance from a decision analysis practitioners perspective while Sriboonchitta (2010) provides a mathematical analysis. Among others, Noyan (2010) and Dentcheva & Ruszczynski (2006) demonstrate how to include stochastic dominance constraints in portfolio optimisation problems.

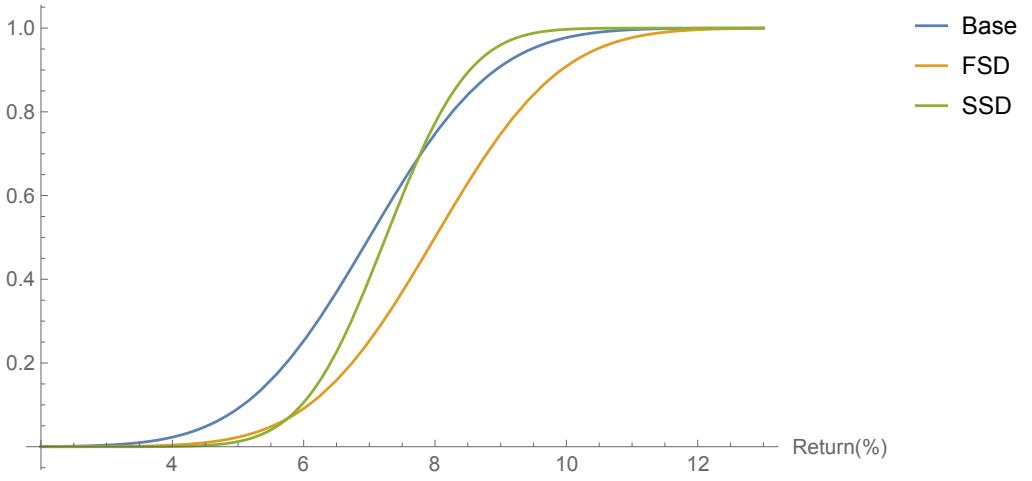



**Figure 42: Stochastic Dominance**

## *7.7 Full Models*

Below are the full models from various sections in the thesis.

### 7.7.1 Generalised Cost Structures

Following adaptation to address generalisation the cost structure for each generation technology to reflect diminishing efficiency in generation, the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint.

Market Equilibrium Conditions

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i(m),t}} + \varphi^{+}_{i(m),r,t}\Big|_{i>0} \geq 0 \quad \perp \quad GEN_{i(m),r,t} \geq 0 \qquad \forall i(m),r,t \tag{7.46}$$

$$\sum_{i(m)} GEN_{i(m),r,t} - \mathrm{L}_{r,t} = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \tag{7.47}$$

$$CAP_{i(m),t} - GEN_{i(m),r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i(m),r,t} \geq 0 \qquad \forall\, i(m)\Big|_{i>0},r,t \tag{7.48}$$

Equilibrium Capacity Choice Conditions

$$\chi_{i(m),t} - \sum_{r<R} \varphi^{+}_{i(m),r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i(m),t}\ free \qquad \forall\, i(m)\Big|_{i>0},t \tag{7.49}$$

$$CAP_{i(m)} - \alpha_{i(m)} CAP_{i} = 0 \quad \perp \quad \chi_{i(m)}\ free \qquad \forall\, i(m)\Big|_{i>0} \tag{7.50}$$

$$\mathrm{FC_i} - \sum_{t} \mathrm{w_t} \sum_{i(m)} \alpha_{i(m)} \chi_{i(m),t} \geq 0 \quad \perp \quad INV_{i} \geq 0 \qquad \forall i > 0 \tag{7.51}$$

Defining Optimal Trade-offs

$$\chi_{i(m),t} - \chi_{j(m),t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i(m), j(m) \neq i(m), t \quad (7.52)$$

$$1 - u^{e}_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \quad (7.53)$$

Selecting Critical Utilisation Levels

$$1 - \psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.54)$$

$$\psi^{0}_{n,t} \sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC_{j,t}} + \psi^{3}_{n,t} z_{j,n\text{-}1,t} + 2\psi^{4}_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j, n > 0, t \quad (7.55)$$

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t} \sum_{i} z_{i,n\text{-}1,t} u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.56)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.57)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC_{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC_{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.58)$$

$$-\sum_{j} z_{j,n\text{-}1,t} z_{j,n,t} + \psi^{5}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.59)$$

$$\sum_{j} {z_{j,n,t}}^{2} - 1 \geq 0 \quad \perp \quad \psi^{4}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.60)$$

$$1 - u^{e}_{n,t} \geq 0 \quad \perp \quad \psi^{5}_{n,t} \geq 0 \qquad \forall n > 0, t \quad (7.61)$$

Ordering Utilisation Levels

$$-\mathrm{r.u_{k,t}} + \phi^{0}_{r,t} + \phi^{f}_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k, r, t \quad (7.62)$$

$$-\mathrm{r}.u^{e}_{n,t} + \phi^{0}_{r,t} + \phi^{e}_{n,t} \geq 0 \quad \perp \quad x^{e}_{n,r,t} \geq 0 \qquad \forall n, r, t \quad (7.63)$$

$$1 - \sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{0}_{r,t} \geq 0 \qquad \forall r, t \quad (7.64)$$

$$1 - \sum_{r} x_{k,r,t} \geq 0 \quad \perp \quad \phi^{f}_{k,t} \geq 0 \qquad \forall k, t \quad (7.65)$$

$$1 - \sum_{r} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{e}_{n,t} \geq 0 \qquad \forall n, t \quad (7.66)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_{k} x_{k,r,t}\mathrm{u_{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} = 0 \qquad \forall r, t \quad (7.67)$$

$$L_{r,t} - \sum_{k} x_{k,r,t}\mathrm{L_{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r, t \quad (7.68)$$

$$L^{e}_{n,t} - \mathrm{L_{0,t}} + \sum_{k} \frac{\mathrm{L_{k\text{-}1,t}} - \mathrm{L_{k,t}}}{\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n, t \quad (7.69)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n,t \qquad (7.70)$$

$$\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < k < K, n, t \qquad (7.71)$$

Initial Conditions and Definitions

$$u^{e}_{0,t} = 0 \qquad \forall t \qquad (7.72)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.73)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (7.74)$$

### 7.7.2 Capacity Inflexibility

Following adaptation to address retirement and mothballing, the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint.

Market Equilibrium Conditions

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} + \varphi^{+}_{i,r,t}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.75)$$

$$\sum_{i} GEN_{i,r,t} - \mathrm{L}_{r,t} = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \qquad (7.76)$$

$$CAP_{i,t} - MBL_{i} + REI_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,r,t} \geq 0 \qquad \forall i > 0, r, t \qquad (7.77)$$

Mothballing and Reinstatement

$$\mathrm{a}^{\mathrm{REI}}_{\mathrm{i,t}} MBL_{i} - REI_{i,t} \geq 0 \quad \perp \quad \chi^{REI}_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (7.78)$$

$$\mathrm{MC}^{\mathrm{REI}}_{\mathrm{i}} - \mathrm{w_t}\chi_{i,t} + \chi^{REI}_{i,t} \geq 0 \quad \perp \quad REI_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (7.79)$$

$$CAP_{i} - MBL_{i} \geq 0 \quad \perp \quad \chi^{MBL}_{i} \geq 0 \qquad \forall i > 0 \qquad (7.80)$$

$$\sum_{t} \mathrm{w_t}\chi_{i,t} - \pi^{MBL}_{i} + \chi^{MBL}_{i} \geq 0 \quad \perp \quad MBL_{i} \geq 0 \qquad \forall i > 0 \qquad (7.81)$$

Capacity Limitations

$$CAP_{i} - CAP^{-}_{i} \geq 0 \quad \perp \quad \chi^{-}_{i} \geq 0 \qquad \forall i > 0 \qquad (7.82)$$

$$CAP^{+}_{i} - CAP_{i} \geq 0 \quad \perp \quad \chi^{+}_{i} \geq 0 \qquad \forall i > 0 \qquad (7.83)$$

Equilibrium Capacity Choice Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi^{+}_{i,r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i > 0, t \qquad (7.84)$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (7.85)$$

$$\left(\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}\right) - \mathrm{FOC}_\mathrm{i} - \left(\chi_i^+ - \chi_i^-\right) + \chi_i^{RET} \geq 0 \quad \perp \quad RET_i \geq 0 \qquad \forall i > 0 \qquad (7.86)$$

$$\mathrm{FC}_\mathrm{i} - \left(\sum_t \mathrm{w}_t \chi_{i,t} + \chi_i^{MBL}\right) + \left(\chi_i^+ - \chi_i^-\right) \geq 0 \quad \perp \quad INV_i \geq 0 \qquad \forall i > 0 \qquad (7.87)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC}_{\mathrm{j,t}} - \mathrm{MC}_{\mathrm{i,t}}) u_{i,j,t}^e + \eta_{i,j,t} \geq 0 \quad \perp \quad u_{i,j,t}^e \geq 0 \qquad \forall i, j \neq i, t \qquad (7.88)$$

$$1 - u_{i,j,t}^e \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \qquad (7.89)$$

Selecting Critical Utilisation Levels

$$1 - \psi_{n,t}^0 \geq 0 \quad \perp \quad u_{n,t}^e \geq 0 \qquad \forall n > 0, t \qquad (7.90)$$

$$\psi_{n,t}^0 \sum_i z_{i,n\text{-}1,t} - \psi_{n,t}^1 + \psi_{n,t}^2 \mathrm{MC}_{\mathrm{j,t}} + \psi_{n,t}^3 z_{j,n\text{-}1,t} + 2\psi_{n,t}^4 z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j, n > 0, t \qquad (7.91)$$

$$u_{n,t}^e - \sum_j z_{j,n,t} \sum_i z_{i,n\text{-}1,t} u_{i,j,t}^e \geq 0 \quad \perp \quad \psi_{n,t}^0 \geq 0 \qquad \forall n > 0, t \qquad (7.92)$$

$$\sum_j z_{j,n,t} \geq 1 \quad \perp \quad \psi_{n,t}^1 \geq 0 \qquad \forall n > 0, t \qquad (7.93)$$

$$\sum_j z_{j,n\text{-}1,t} \mathrm{MC}_{\mathrm{j,t}} - \sum_j z_{j,n,t} \mathrm{MC}_{\mathrm{j,t}} \geq 0 \quad \perp \quad \psi_{n,t}^2 \geq 0 \qquad \forall n > 0, t \qquad (7.94)$$

$$-\sum_j z_{j,n\text{-}1,t} z_{j,n,t} + \psi_{n,t}^5 \geq 0 \quad \perp \quad \psi_{n,t}^3 \geq 0 \qquad \forall n > 0, t \qquad (7.95)$$

$$\sum_j {z_{j,n,t}}^2 - 1 \geq 0 \quad \perp \quad \psi_{n,t}^4 \geq 0 \qquad \forall n > 0, t \qquad (7.96)$$

$$1 - u_{n,t}^e \geq 0 \quad \perp \quad \psi_{n,t}^5 \geq 0 \qquad \forall n > 0, t \qquad (7.97)$$

Ordering Utilisation Levels

$$-\mathrm{r.u}_{\mathrm{k,t}} + \phi_{r,t}^0 + \phi_{k,t}^f \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k, r, t \qquad (7.98)$$

$$-\mathrm{r}.u_{n,t}^e + \phi_{r,t}^0 + \phi_{n,t}^e \geq 0 \quad \perp \quad x_{n,r,t}^e \geq 0 \qquad \forall n, r, t \qquad (7.99)$$

$$1 - \sum_k x_{k,r,t} - \sum_n x_{n,r,t}^e \geq 0 \quad \perp \quad \phi_{r,t}^0 \geq 0 \qquad \forall r, t \qquad (7.100)$$

$$1 - \sum_r x_{k,r,t} \geq 0 \quad \perp \quad \phi_{k,t}^f \geq 0 \qquad \forall k, t \qquad (7.101)$$

$$1 - \sum_r x_{n,r,t}^e \geq 0 \quad \perp \quad \phi_{n,t}^e \geq 0 \qquad \forall n, t \qquad (7.102)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_{k} x_{k,r,t} \mathrm{u}_{\mathrm{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} = 0 \qquad \forall r,t \qquad (7.103)$$

$$L_{r,t} - \sum_{k} x_{k,r,t} \mathrm{L}_{\mathrm{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r,t \qquad (7.104)$$

$$L^{e}_{n,t} - \mathrm{L}_{0,\mathrm{t}} + \sum_{k} \frac{\mathrm{L}_{\mathrm{k\text{-}1,t}} - \mathrm{L}_{\mathrm{k,t}}}{\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n,t \qquad (7.105)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n,t \qquad (7.106)$$

$$\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < k < K, n, t \qquad (7.107)$$

Initial Conditions and Definitions

$$u^{e}_{0,t} = 0 \qquad \forall t \qquad (7.108)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.109)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (7.110)$$

$$CAP_i = CAP_i^0 + INV_i - RET_i \qquad \forall i > 0 \qquad (7.111)$$

$$\pi^{MBL}_{i} = \mathrm{FOC_i} - \mathrm{FOC_i^{MBL}} + \sum_{t} \mathrm{w_t a_{i,t}^{REI}} \chi^{REI}_{i,t} - \mathrm{FC^{MBL}} \qquad \forall i > 0 \qquad (7.112)$$

### 7.7.3 Energy Limits

Following adaptation to incorporate energy limits, the optimal trade-off, market clearing and equilibrium capacity model is presented below. The problem is stated in straight capacity terms, but can be adapted to include investment and retirement explicitly. By inspection, the system is square with one constraint for each variable, and one variable for each constraint.

Market Equilibrium Conditions

$$-\lambda_{r,t} + \varepsilon_{i,t} + \varphi^{+}_{i,r,t}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.113)$$

$$\sum_{i} GEN_{i,r,t} - L_{r,t} = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \qquad (7.114)$$

$$CAP_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi^{+}_{i,r,t} \geq 0 \qquad \forall i > 0, r, t \qquad (7.115)$$

Inflow Definition

$$\mathrm{MC_{i,t}} - \varepsilon_{i,t} + \eta_{i,t} \geq 0 \quad \perp \quad INF_i \geq 0 \qquad \forall i > 0, t \qquad (7.116)$$

$$\mathrm{INF^{+}_{i,t}} - INF_i \geq 0 \quad \perp \quad \eta_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (7.117)$$

Storage Restrictions

$$STOR_{i,t} - \mathrm{STOR}^{-}_{\mathrm{i,t}} \geq 0 \quad \perp \quad \gamma^{-}_{i,t} \geq 0 \qquad \forall i>0,t \qquad (7.118)$$

$$\mathrm{STOR}^{+}_{\mathrm{i,t}} - STOR_{i,t} \geq 0 \quad \perp \quad \gamma^{+}_{i,t} \geq 0 \qquad \forall i>0,t \qquad (7.119)$$

$$\mathrm{STOR}_{\mathrm{i,0}}\Big|_{t=1} + STOR_{i,t-1}\Big|_{t>1} + \mathrm{INF}_{\mathrm{i,t}} - REL_{i,t} - STOR_{i,t} = 0 \quad \perp \quad \gamma_{i,t}\ free \qquad \forall i>0,t \qquad (7.120)$$

Energy Management

$$REL_{i,t} - \frac{1}{2}\sum_{r=0}^{R-1}\left[\left(GEN_{i,r,t} - GEN_{i,r+1,t}\right)\left(u_{r+1,t} + u_{r,t}\right)\right] \geq 0 \quad \perp \quad \varepsilon_{i,t} \geq 0 \qquad \forall i>0,t \qquad (7.121)$$

$$\gamma_{i,t} - \varepsilon_{i,t} \geq 0 \quad \perp \quad REL_{i,t} \geq 0 \qquad \forall i>0,t \qquad (7.122)$$

$$\gamma_{i,t} - \gamma_{i,t+1}\Big|_{t<T} - V_i'\left(STOR_{i,T}\right)\Big|_{t=T} + \gamma^{+}_{i,t} - \gamma^{-}_{i,t} = 0 \quad \perp \quad STOR_{i,t}\ free \qquad \forall i>0,t \qquad (7.123)$$

Equilibrium Capacity Choice Conditions

$$\chi_{i,t} - \sum_{r<R}\varphi^{+}_{i,r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i,t \qquad (7.124)$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i>0,t \qquad (7.125)$$

$$\mathrm{FC}_{\mathrm{i}} - \sum_{t}\mathrm{w}_{\mathrm{t}}\chi_{i,t} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i>0 \qquad (7.126)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\varepsilon_{j,t} - \varepsilon_{i,t})u^{e}_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^{e}_{i,j,t} \geq 0 \qquad \forall i,j\neq i,t \qquad (7.127)$$

$$1 - u^{e}_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i,j\neq i,t \qquad (7.128)$$

Selecting Critical Utilisation Levels

$$1 - \psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.129)$$

$$\psi^{0}_{n,t}\sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC}_{\mathrm{j,t}} + \psi^{3}_{n,t}z_{j,n\text{-}1,t} + 2\psi^{4}_{n,t}z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j,n>0,t \qquad (7.130)$$

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t}\sum_{i} z_{i,n\text{-}1,t}u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.131)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.132)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC}_{\mathrm{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC}_{\mathrm{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.133)$$

$$-\sum_{j} z_{j,n\text{-}1,t}z_{j,n,t} + \psi^{5}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.134)$$

$$\sum_j z_{j,n,t}{}^2 - 1 \geq 0 \quad \perp \quad \psi^4_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.135)$$

$$1 - u^e_{n,t} \geq 0 \quad \perp \quad \psi^5_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.136)$$

Ordering Utilisation Levels

$$-\mathrm{r.u_{k,t}} + \phi^0_{r,t} + \phi^f_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k, r, t \qquad (7.137)$$

$$-\mathrm{r}.u^e_{n,t} + \phi^0_{r,t} + \phi^e_{n,t} \geq 0 \quad \perp \quad x^e_{n,r,t} \geq 0 \qquad \forall n, r, t \qquad (7.138)$$

$$1 - \sum_k x_{k,r,t} - \sum_n x^e_{n,r,t} \geq 0 \quad \perp \quad \phi^0_{r,t} \geq 0 \qquad \forall r, t \qquad (7.139)$$

$$1 - \sum_r x_{k,r,t} \geq 0 \quad \perp \quad \phi^f_{k,t} \geq 0 \qquad \forall k, t \qquad (7.140)$$

$$1 - \sum_r x^e_{n,r,t} \geq 0 \quad \perp \quad \phi^e_{n,t} \geq 0 \qquad \forall n, t \qquad (7.141)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_k x_{k,r,t} \mathrm{u_{k,t}} + \sum_n x^e_{n,r,t} u^e_{n,t} = 0 \qquad \forall r, t \qquad (7.142)$$

$$L_{r,t} - \sum_k x_{k,r,t} \mathrm{L_{k,t}} + \sum_n x^e_{n,r,t} L^e_{n,t} = 0 \qquad \forall r, t \qquad (7.143)$$

$$L^e_{n,t} - \mathrm{L_{0,t}} + \sum_k \frac{\mathrm{L_{k\text{-}1,t}} - \mathrm{L_{k,t}}}{\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n, t \qquad (7.144)$$

$$\sum_k u^{part}_{k,n,t} - u^e_{n,t} = 0 \qquad \forall n, t \qquad (7.145)$$

$$\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < k < K, n, t \qquad (7.146)$$

Initial Conditions and Definitions

$$u^e_{0,t} = 0 \qquad \forall t \qquad (7.147)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.148)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (7.149)$$

### 7.7.4 Configurable Technologies

The introduction of configurable technologies involves significant modification to the model. Following adaptation, the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint, but we note for the reader the set of technologies i, is comprised of a set C

which indexes conventional technologies and a set i(r), which is an operating range specific index of configurable technologies.

Market Equilibrium Conditions

$$-\lambda_r + \mathrm{MC}_{\mathrm{i}}^{+} + \varphi_{i,r}^{+} \geq 0 \quad \perp \quad GEN_i \geq 0 \qquad \forall i \notin C, r \qquad (7.150)$$

$$-\lambda_r + \left( \mathrm{MC}_{\mathrm{i(r)}}^{+} - \frac{1}{2}\left[ \frac{\left(\mathrm{MC}_{\mathrm{i(r)}}^{+} - \mathrm{MC}_{\mathrm{i(r)}}^{-}\right)^2}{\mathrm{FC}_{\mathrm{i(r)}}^{+} - \mathrm{FC}_{\mathrm{i(r)}}^{-}} \right] u_r \right) + \varphi_{i(r),r}^{+} \geq 0 \quad \perp \quad GEN_{i(r)} \geq 0 \qquad \forall i(r), r \qquad (7.151)$$

$$\sum_{i \notin C} GEN_{i,r} + \sum_{i(r)} GEN_{i(r)} - L_r = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r, t \qquad (7.152)$$

$$CAP_i - GEN_{i,r} \geq 0 \quad \perp \quad \varphi_{i,r}^{+} \geq 0 \qquad \forall i > 0, i \notin C, r \qquad (7.153)$$

$$CAP_{i(r)} - GEN_{i(r),r} \geq 0 \quad \perp \quad \varphi_{i(r),r}^{+} \geq 0 \qquad \forall i(r), r \qquad (7.154)$$

Equilibrium Capacity Choice Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi_{i,r+1,t}^{+} \left( u_{r+1,t} - u_{r,t} \right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i, t \qquad (7.155)$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0, t \qquad (7.156)$$

$$\mathrm{FC}_{\mathrm{i}} - \sum_{r<R} \left[ \varphi_{i,r+1}^{+} + \frac{1}{2}\left( \varphi_{i,r}^{+,\varepsilon} + \varphi_{i,r+1}^{+} \right) \left( \frac{u_{r+1} - u_r}{u_{r+1} - u_r^{\varepsilon}} \right) \right] \left( \mathrm{u}_{r+1} - \mathrm{u}_r \right) \geq 0 \perp CAP_i \geq 0 \qquad \forall i \notin C \qquad (7.157)$$

$$\mathrm{FC}_{\mathrm{i(r)}} - \sum_{r<R} \left[ \varphi_{i(r),r+1}^{+} + \frac{1}{2}\left( \varphi_{i(r),r}^{+,\varepsilon} + \varphi_{i(r),r+1}^{+} \right) \left( \frac{u_{r+1} - u_r}{u_{r+1} - u_r^{\varepsilon}} \right) \right] \left( \mathrm{u}_{r+1} - \mathrm{u}_r \right) \geq 0 \perp CAP_{i(r)} \geq 0 \qquad \forall i(r) \qquad (7.158)$$

Defining Optimal Trade-offs:

$$u_{i,j,v}^{e} - \left( 1 - \frac{\mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} + \mathrm{a}_{\mathrm{j}}^{\mathrm{cfg}}}{\mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} + \mathrm{a}_{\mathrm{j}}^{\mathrm{cfg}} + \varepsilon} \right) \frac{-c}{b} - \left( \frac{\mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} + \mathrm{a}_{\mathrm{j}}^{\mathrm{cfg}}}{\mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} + \mathrm{a}_{\mathrm{j}}^{\mathrm{cfg}} + \varepsilon} \right) \left[ \frac{-b_{i,j} + \sqrt{\zeta_{i,j}^{1} + \frac{\zeta_{i,j}^{1}}{\varepsilon^{1}}}}{2\left( \zeta_{i,j}^{3} + \zeta_{i,j}^{4} + \left( 1 - \frac{\zeta_{i,j}^{3} + \zeta_{i,j}^{4}}{\zeta_{i,j}^{3} + \zeta_{i,j}^{4} + \varepsilon^{2}} \right) \varepsilon^{2} \right)} \right] + \eta_{i,j,v} \geq 0$$

$$\perp \quad u_{i,j,v}^{e} \geq 0 \qquad \forall i, j \neq i, v \qquad (7.159)$$

$$\mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} \left[ \frac{2\left( \mathrm{FC}_{\mathrm{i}}^{+} - \mathrm{FC}_{\mathrm{i}}^{-} \right)}{\left( \mathrm{MC}_{\mathrm{i}}^{+} - \mathrm{MC}_{\mathrm{i}}^{-} \right)} \right] + \left( 1 - \mathrm{a}_{\mathrm{i}}^{\mathrm{cfg}} \right) - u_{i,j}^{e} \geq 0 \quad \perp \quad \eta_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (7.160)$$

$$\zeta_{i,j}^{1} - {b_{i,j}}^{2} + 4 a_{i,j} c_{i,j} \geq 0 \quad \perp \quad \zeta_{i,j}^{1} \geq 0 \qquad \forall i, j \neq i \qquad (7.161)$$

$${b_{i,j}}^{2} - 4 a_{i,j} c_{i,j} - \zeta_{i,j}^{1} + \zeta_{i,j}^{2} \geq 0 \quad \perp \quad \zeta_{i,j}^{2} \geq 0 \qquad \forall i, j \neq i \qquad (7.162)$$

$$\zeta^3_{i,j} - a_{i,j} \geq 0 \quad \perp \quad \zeta^3_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (7.163)$$

$$a_{i,j} - \zeta^3_{i,j} + \zeta^4_{i,j} \geq 0 \quad \perp \quad \zeta^4_{i,j} \geq 0 \qquad \forall i, j \neq i \qquad (7.164)$$

Selecting Critical Utilisation Levels

$$1 - \psi^0_{n,t} \geq 0 \quad \perp \quad u^e_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.165)$$

$$\psi^0_{n,t} \sum_i z_{i,n\text{-}1,t} - \psi^1_{n,t} + \psi^2_{n,t} \mathrm{MC}_{\mathrm{j,t}} + \psi^3_{n,t} z_{j,n\text{-}1,t} + 2\psi^4_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j, n > 0, t \qquad (7.166)$$

$$u^e_n - \sum_j z_{j,n,v} \sum_{i,v} \mathrm{z}_{\mathrm{i,n\text{-}1,v}} \mathrm{u}^{\mathrm{e}}_{\mathrm{i,j,v}} = 0 \qquad : \psi^0_n \qquad \forall n > 0 \qquad (7.167)$$

$$\sum_{j,v} z_{j,n,v} - 1 \geq 0 \qquad : \psi^1_n \qquad \forall n \qquad (7.168)$$

$$\sum_j z_{j,n\text{-}1,t} \mathrm{MC}_{\mathrm{j,t}} - \sum_j z_{j,n,t} \mathrm{MC}_{\mathrm{j,t}} \geq 0 \quad \perp \quad \psi^2_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.169)$$

$$-\sum_j z_{j,n\text{-}1,t} z_{j,n,t} + \psi^5_{n,t} \geq 0 \quad \perp \quad \psi^3_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.170)$$

$$\sum_{j,v} z_{j,n,v}{}^2 - 1 \geq 0 \qquad : \psi^5_n \qquad \forall n \qquad (7.171)$$

$$1 - u^e_{n,t} \geq 0 \quad \perp \quad \psi^5_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.172)$$

Ordering Utilisation Levels

$$-\mathrm{r}.\mathrm{u}_{\mathrm{k,t}} + \phi^0_{r,t} + \phi^f_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k, r, t \qquad (7.173)$$

$$-\mathrm{r}.u^e_{n,t} + \phi^0_{r,t} + \phi^e_{n,t} \geq 0 \quad \perp \quad x^e_{n,r,t} \geq 0 \qquad \forall n, r, t \qquad (7.174)$$

$$1 - \sum_k x_{k,r,t} - \sum_n x^e_{n,r,t} \geq 0 \quad \perp \quad \phi^0_{r,t} \geq 0 \qquad \forall r, t \qquad (7.175)$$

$$1 - \sum_r x_{k,r,t} \geq 0 \quad \perp \quad \phi^f_{k,t} \geq 0 \qquad \forall k, t \qquad (7.176)$$

$$1 - \sum_r x^e_{n,r,t} \geq 0 \quad \perp \quad \phi^e_{n,t} \geq 0 \qquad \forall n, t \qquad (7.177)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_k x_{k,r,t} \mathrm{u}_{\mathrm{k,t}} + \sum_n x^e_{n,r,t} u^e_{n,t} = 0 \qquad \forall r, t \qquad (7.178)$$

$$L_{r,t} - \sum_k x_{k,r,t} \mathrm{L}_{\mathrm{k,t}} + \sum_n x^e_{n,r,t} L^e_{n,t} = 0 \qquad \forall r, t \qquad (7.179)$$

$$L^e_{n,t} - \mathrm{L}_{\mathrm{0,t}} + \sum_k \frac{\mathrm{L}_{\mathrm{k\text{-}1,t}} - \mathrm{L}_{\mathrm{k,t}}}{\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n, t \qquad (7.180)$$

$$\sum_{k} u_{k,n,t}^{part} - u_{n,t}^{e} = 0 \qquad \forall n,t \qquad (7.181)$$

$$\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}} - u_{k,n,t}^{part} \geq 0 \quad \perp \quad u_{k+1,n,t}^{part} \geq 0 \qquad \forall 0 < k < K, n, t \qquad (7.182)$$

Initial Conditions and Definitions

$$u_{0,t}^{e} = 0 \qquad \forall t \qquad (7.183)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.184)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (7.185)$$

$$\mathrm{FC_{i(r)}} = \mathrm{FC^{-}_{i(r)}} + \frac{1}{4}\left(\frac{\left(\mathrm{MC_i^{+}} - \mathrm{MC_i^{-}}\right)^2}{\mathrm{FC_i^{+}} - \mathrm{FC_i^{-}}}\right) u_r^2 \qquad \forall i \qquad (7.186)$$

### 7.7.5 Long Term Demand Response

Following adaptation to account for load response the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint, but we draw the reader's attention to the indexation of the capacity constraints. In particular, the investment constraint does not include the demand response technology, whose capacity level is defined directly.

Market Equilibrium Conditions:

$$-\lambda_{r,t} + \mathrm{MC_{i,t}} + \varphi_{i,r,t}^{+}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.187)$$

$$\sum_{i} GEN_{i,r,t} - L_{r,t} = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \qquad (7.188)$$

$$CAP_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi_{i,r,t}^{+} \geq 0 \qquad \forall 0 < i,r,t \qquad (7.189)$$

Optimal Capacity Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi_{i,r+1,t}^{+}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i,t \qquad (7.190)$$

$$CAP_{i} - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall 0 < i,t \qquad (7.191)$$

$$\mathrm{FC_i} - \sum_{t} \mathrm{w_t} \chi_{i,t} \geq 0 \quad \perp \quad CAP_{i} \geq 0 \qquad \forall 0 < i < I+1 \qquad (7.192)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}}) u_{i,j,t}^{e} + \eta_{i,j,t} \geq 0 \quad \perp \quad u_{i,j,t}^{e} \geq 0 \qquad \forall i, j \neq i, t \qquad (7.193)$$

$$1 - u_{i,j,t}^{e} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \qquad (7.194)$$

Selecting Critical Utilisation Levels

$$1-\psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.195)$$

$$\psi^{0}_{n,t}\sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC_{j,t}} + \psi^{3}_{n,t} z_{j,n\text{-}1,t} + 2\psi^{4}_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j,n>0,t \qquad (7.196)$$

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t} \sum_{i} z_{i,n\text{-}1,t} u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.197)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.198)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC_{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC_{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.199)$$

$$-\sum_{j} z_{j,n\text{-}1,t} z_{j,n,t} + \psi^{5}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.200)$$

$$\sum_{j} {z_{j,n,t}}^{2} - 1 \geq 0 \quad \perp \quad \psi^{4}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.201)$$

$$1-u^{e}_{n,t} \geq 0 \quad \perp \quad \psi^{5}_{n,t} \geq 0 \qquad \forall n>0,t \qquad (7.202)$$

Ordering Utilisation Levels

$$-\mathrm{r.u_{k,t}} + \phi^{0}_{r,t} + \phi^{f}_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k,r,t \qquad (7.203)$$

$$-\mathrm{r}.u^{e}_{n,t} + \phi^{0}_{r,t} + \phi^{e}_{n,t} \geq 0 \quad \perp \quad x^{e}_{n,r,t} \geq 0 \qquad \forall n,r,t \qquad (7.204)$$

$$1-\sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{0}_{r,t} \geq 0 \qquad \forall r,t \qquad (7.205)$$

$$1-\sum_{r} x_{k,r,t} \geq 0 \quad \perp \quad \phi^{f}_{k,t} \geq 0 \qquad \forall k,t \qquad (7.206)$$

$$1-\sum_{r} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{e}_{n,t} \geq 0 \qquad \forall n,t \qquad (7.207)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_{k} x_{k,r,t}\mathrm{u_{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} = 0 \qquad \forall r,t \qquad (7.208)$$

$$L_{r,t} - \sum_{k} x_{k,r,t}\mathrm{L_{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r,t \qquad (7.209)$$

$$L^{e}_{n,t} - \mathrm{L_{0,t}} + \sum_{k} \frac{\mathrm{L_{k\text{-}1,t}} - \mathrm{L_{k,t}}}{\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n,t \qquad (7.210)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n,t \qquad (7.211)$$

$$\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0<k<K,n,t \qquad (7.212)$$

Resolving Ambiguity

$$a^0_{i,r,t} + \left( \sum_{n=1}^{N} z_{i,n-1,t} u^e_{n,t} - u_{r+1,t} \right) \geq 0 \quad \perp \quad a^1_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.213)$$

$$a^1_{i,r,t} + GEN_{i,r,t} \geq 0 \quad \perp \quad a^0_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.214)$$

Initial Conditions & Definitions:

$$u^e_{0,t} = 0 \qquad \forall t \qquad (7.215)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.216)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0,t \qquad (7.217)$$

$$CAP_{I+1,r,t} - \mathrm{a}_{\mathrm{t}}^{\mathrm{shift}} \lambda_{r,t} - \left( \mathrm{a}_{\mathrm{t}}^{\mathrm{cont}} - \mathrm{a}_{\mathrm{t}}^{\mathrm{shift}} \right) \frac{\sum_{r,t} \lambda_{r,t} L_{r,t}}{\sum_{r,t} L_{r,t}} = 0 \qquad \forall r,t \qquad (7.218)$$

### 7.7.6 Reliability Model

Following adaptation to account for reliability issues, the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint.

Market Equilibrium Conditions:

$$-\lambda_{r,t} + \mathrm{MC}_{\mathrm{i,t}} + \varphi^+_{i,r,t}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.219)$$

$$\sum_i GEN_{i,r,t} - \left( L_{r,t} + OUT_{r,t} \right) = 0 \quad \perp \quad \lambda_{r,t}\ free \qquad \forall r,t \qquad (7.220)$$

$$CAP_{i,t} - GEN_{i,r,t} \geq 0 \quad \perp \quad \varphi^+_{i,r,t} \geq 0 \qquad \forall i > 0,r,t \qquad (7.221)$$

$$OUT_{i,r,t} + \rho_i CAP_i - GEN_{i,r,t} \geq 0 \quad \perp \quad OUT_{i,r,t} \geq 0 \qquad \forall i,r,t \qquad (7.222)$$

Optimal Capacity Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi^+_{i,r+1,t} \left( u_{r+1,t} - u_{r,t} \right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i,t \qquad (7.223)$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0,t \qquad (7.224)$$

$$\mathrm{FC_i} - \rho_i \sum_t \mathrm{w_t} \chi_{i,t} \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i > 0 \qquad (7.225)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}}) u^e_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^e_{i,j,t} \geq 0 \qquad \forall i, j \neq i,t \qquad (7.226)$$

$$1-u^{e}_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i, j \neq i, t \qquad (7.227)$$

Selecting Critical Utilisation Levels

$$1-\psi^{0}_{n,t} \geq 0 \quad \perp \quad u^{e}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.228)$$

$$\psi^{0}_{n,t}\sum_{i} z_{i,n\text{-}1,t} - \psi^{1}_{n,t} + \psi^{2}_{n,t}\mathrm{MC_{j,t}} + \psi^{3}_{n,t} z_{j,n\text{-}1,t} + 2\psi^{4}_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j, n > 0, t \qquad (7.229)$$

$$u^{e}_{n,t} - \sum_{j} z_{j,n,t} \sum_{i} z_{i,n\text{-}1,t} u^{e}_{i,j,t} \geq 0 \quad \perp \quad \psi^{0}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.230)$$

$$\sum_{j} z_{j,n,t} \geq 1 \quad \perp \quad \psi^{1}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.231)$$

$$\sum_{j} z_{j,n\text{-}1,t}\mathrm{MC_{j,t}} - \sum_{j} z_{j,n,t}\mathrm{MC_{j,t}} \geq 0 \quad \perp \quad \psi^{2}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.232)$$

$$-\sum_{j} z_{j,n\text{-}1,t} z_{j,n,t} + \psi^{5}_{n,t} \geq 0 \quad \perp \quad \psi^{3}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.233)$$

$$\sum_{j} {z_{j,n,t}}^{2} - 1 \geq 0 \quad \perp \quad \psi^{4}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.234)$$

$$1-u^{e}_{n,t} \geq 0 \quad \perp \quad \psi^{5}_{n,t} \geq 0 \qquad \forall n > 0, t \qquad (7.235)$$

Ordering Utilisation Levels

$$-\mathrm{r.u_{k,t}} + \phi^{0}_{r,t} + \phi^{f}_{k,t} \geq 0 \quad \perp \quad x_{k,r,t} \geq 0 \qquad \forall k, r, t \qquad (7.236)$$

$$-\mathrm{r}.u^{e}_{n,t} + \phi^{0}_{r,t} + \phi^{e}_{n,t} \geq 0 \quad \perp \quad x^{e}_{n,r,t} \geq 0 \qquad \forall n, r, t \qquad (7.237)$$

$$1-\sum_{k} x_{k,r,t} - \sum_{n} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{0}_{r,t} \geq 0 \qquad \forall r, t \qquad (7.238)$$

$$1-\sum_{r} x_{k,r,t} \geq 0 \quad \perp \quad \phi^{f}_{k,t} \geq 0 \qquad \forall k, t \qquad (7.239)$$

$$1-\sum_{r} x^{e}_{n,r,t} \geq 0 \quad \perp \quad \phi^{e}_{n,t} \geq 0 \qquad \forall n, t \qquad (7.240)$$

Defining Variables at Critical Utilisation Levels

$$u_{r,t} - \sum_{k} x_{k,r,t}\mathrm{u_{k,t}} + \sum_{n} x^{e}_{n,r,t} u^{e}_{n,t} = 0 \qquad \forall r, t \qquad (7.241)$$

$$L_{r,t} - \sum_{k} x_{k,r,t}\mathrm{L_{k,t}} + \sum_{n} x^{e}_{n,r,t} L^{e}_{n,t} = 0 \qquad \forall r, t \qquad (7.242)$$

$$L^{e}_{n,t} - \mathrm{L_{0,t}} + \sum_{k} \frac{\mathrm{L_{k\text{-}1,t}} - \mathrm{L_{k,t}}}{\mathrm{u_{k,t}} - \mathrm{u_{k\text{-}1,t}}} u^{part}_{k,n,t} = 0 \qquad \forall n, t \qquad (7.243)$$

$$\sum_{k} u^{part}_{k,n,t} - u^{e}_{n,t} = 0 \qquad \forall n, t \qquad (7.244)$$

$$\mathrm{u}_{\mathrm{k,t}} - \mathrm{u}_{\mathrm{k\text{-}1,t}} - u^{part}_{k,n,t} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < k < K, n, t \quad (7.245)$$

Initial Conditions & Definitions:

$$u^{e}_{0,t} = 0 \qquad \forall t \quad (7.246)$$

$$z_{0,0,t} = 1 \qquad \forall t \quad (7.247)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \quad (7.248)$$

$$OUT_{r,t} = \sum_{i} OUT_{i,r,t} \qquad \forall r, t \quad (7.249)$$

### 7.7.7 Intermittent Generation Model

Following adaptation to account for intermittent generation, the optimal trade-off, market clearing and equilibrium capacity model is presented below. By inspection, the system is square with one constraint for each variable, and one variable for each constraint. We draw the reader's attention to the significant number of definitions. These are included only to ensure that the expressions in the rest of the model are readily interpretable.

Market Equilibrium Conditions:

$$-\lambda_r + \mathrm{MC_i} + \varphi^{+}_{i,r}\big|_{i>0} \geq 0 \quad \perp \quad GEN_{i,r} \geq 0 \qquad \forall i \notin INT, t \quad (7.250)$$

$$\sum_{i \notin INT} GEN_{i,r} - NL_r - SPL_r = 0 \quad \perp \quad \lambda_r \; free \qquad \forall r \quad (7.251)$$

$$\lambda_r \geq 0 \quad \perp \; SPL_r \geq 0 \qquad \forall r \quad (7.252)$$

$$CAP_i - GEN_{i,r} \geq 0 \quad \perp \quad \varphi^{+}_{i,r} \geq 0 \qquad \forall i > 0, i \notin INT, r \quad (7.253)$$

Constructing the Net LDC

$$NL_h - NL_{h-1} + \gamma^1_h \geq 0 \quad \perp \quad \gamma^1_h \geq 0 \qquad \forall h > 0 \quad (7.254)$$

$$NL_{h-1} - NL_h + \gamma^2_h \geq 0 \quad \perp \quad \gamma^2_h \geq 0 \qquad \forall h > 0 \quad (7.255)$$

$$u^{CLP}_{h^*,h} - \frac{\left(\left(1 - \frac{\gamma^1_h}{NL_{h-1} - NL_h}\right) NL_h + \left(1 - \frac{\gamma^2_h}{NL_h - NL_{h-1}}\right) NL_{h-1} - NL^{CLP}_{h^*}\right)}{\gamma^1_h + \gamma^2_h} \left(t_h - t_{h-1}\right) + \gamma^3_{h^*,h} \geq 0 \perp u^{CLP}_{h^*,h} \geq 0$$

$$\forall h^*, h > 0 \quad (7.256)$$

$$t_h - t_{h-1} - u^{CLP}_{h^*,h} \geq 0 \quad \perp \quad \gamma^3_{h^*,h} \geq 0 \qquad \forall h^*, h > 0 \quad (7.257)$$

Optimal Capacity Conditions

$$\chi_{i,t} - \sum_{r<R} \varphi^{+}_{i,r+1,t}\left(u_{r+1,t} - u_{r,t}\right) = 0 \quad \perp \quad CAP_{i,t}\ free \qquad \forall i,t \qquad (7.258)$$

$$CAP_i - CAP_{i,t} \geq 0 \quad \perp \quad \chi_{i,t} \geq 0 \qquad \forall i > 0,t \qquad (7.259)$$

$$\mathrm{FC_i} - \sum_{h^*} \int_{t_h^*}^{t_{h^*+1}} REV_{h^*}(t)\,dt \geq 0 \quad \perp \quad CAP_i \geq 0 \qquad \forall i \in INT \qquad (7.260)$$

Defining Optimal Trade-offs

$$\chi_{i,t} - \chi_{j,t} - (\mathrm{MC_{j,t}} - \mathrm{MC_{i,t}})u^e_{i,j,t} + \eta_{i,j,t} \geq 0 \quad \perp \quad u^e_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (7.261)$$

$$1 - u^e_{i,j,t} \geq 0 \quad \perp \quad \eta_{i,j,t} \geq 0 \qquad \forall i,j \neq i,t \qquad (7.262)$$

Selecting Critical Utilisation Levels

$$1 - \psi^0_{n,t} \geq 0 \quad \perp \quad u^e_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.263)$$

$$\psi^0_{n,t} \sum_i z_{i,n\text{-}1,t} - \psi^1_{n,t} + \psi^2_{n,t}\mathrm{MC_{j,t}} + \psi^3_{n,t} z_{j,n\text{-}1,t} + 2\psi^4_{n,t} z_{j,n,t} \geq 0 \quad \perp \quad z_{j,n,t} \geq 0 \qquad \forall j,n > 0,t \qquad (7.264)$$

$$u^e_{n,t} - \sum_j z_{j,n,t} \sum_i z_{i,n\text{-}1,t} u^e_{i,j,t} \geq 0 \quad \perp \quad \psi^0_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.265)$$

$$\sum_j z_{j,n,t} \geq 1 \quad \perp \quad \psi^1_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.266)$$

$$\sum_j z_{j,n\text{-}1,t}\mathrm{MC_{j,t}} - \sum_j z_{j,n,t}\mathrm{MC_{j,t}} \geq 0 \quad \perp \quad \psi^2_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.267)$$

$$-\sum_j z_{j,n\text{-}1,t} z_{j,n,t} + \psi^5_{n,t} \geq 0 \quad \perp \quad \psi^3_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.268)$$

$$\sum_j {z_{j,n,t}}^2 - 1 \geq 0 \quad \perp \quad \psi^4_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.269)$$

$$1 - u^e_{n,t} \geq 0 \quad \perp \quad \psi^5_{n,t} \geq 0 \qquad \forall n > 0,t \qquad (7.270)$$

Ordering Utilisation Levels

$$-\mathrm{r}.u^{CLP}_{h^*} + \phi^0_r + \phi^{CLP}_{h^*} \geq 0 \quad \perp \quad x^{CLP}_{h^*,r} \geq 0 \qquad \forall h^*,r \qquad (7.271)$$

$$-\mathrm{r}.u^e_n + \phi^0_r + \phi^e_n \geq 0 \quad \perp \quad x^e_{n,r} \geq 0 \qquad \forall n,r \qquad (7.272)$$

$$1 - \sum_{h^*} x^{CLP}_{h^*,r} - \sum_n x^e_{n,r} \geq 0 \quad \perp \quad \phi^0_r \geq 0 \qquad \forall r \qquad (7.273)$$

$$1 - \sum_r x^{CLP}_{h^*,r} \geq 0 \quad \perp \quad \phi^{CLP}_{h^*} \geq 0 \qquad \forall h^* \qquad (7.274)$$

$$1 - \sum_r x^e_{n,r} \geq 0 \quad \perp \quad \phi^e_n \geq 0 \qquad \forall n \qquad (7.275)$$

Defining Variables at Critical Utilisation Levels

$$u_r - \sum_{h^*} x^{CLP}_{h^*,r} u^{CLP}_{h^*} + \sum_n x^e_{n,r} u^e_n = 0 \qquad \forall r \qquad (7.276)$$

$$NL_r - \sum_{h^*} x^{CLP}_{h^*,r} NL_{h^*} + \sum_n x^e_{n,r} NL^e_n = 0 \qquad \forall r \qquad (7.277)$$

$$NL^e_n - NL_0 + \sum_{h^*} \frac{NL_{h^*-1} - NL_{h^*}}{u^{CLP}_{h^*} - u^{CLP}_{h^*-1}} u^{part}_{h^*,n} = 0 \qquad \forall n \qquad (7.278)$$

$$\sum_{h^*} u^{part}_{h^*,n} - u^e_{n,t} = 0 \qquad \forall n \qquad (7.279)$$

$$u^{CLP}_{h^*} - u^{CLP}_{h^*-1} - u^{part}_{h^*,n} \geq 0 \quad \perp \quad u^{part}_{k+1,n,t} \geq 0 \qquad \forall 0 < h^* < H, n \qquad (7.280)$$

Initial Conditions:

$$u^e_{0,t} = 0 \qquad \forall t \qquad (7.281)$$

$$z_{0,0,t} = 1 \qquad \forall t \qquad (7.282)$$

$$z_{0,i,t} = 0 \qquad \forall i > 0, t \qquad (7.283)$$

$$GEN^{INT}_h = \sum_{i \in INT} \mathrm{ICF_{i,h}} CAP_i \qquad \forall h \qquad (7.284)$$

$$NL^{CLP}_h = \mathrm{L^{CLP}_h} - \sum_{i \in INT} \mathrm{ICF_{i,h}} CAP_i \qquad \forall h \qquad (7.285)$$

$$u^{CLP}_{h^*} = \sum_{h=1}^{H} u^{CLP}_{h^*,h} \qquad \forall h^* \qquad (7.286)$$

$$NL_{h^*} = NL^{CLP}_{h^*} S\left(u^{CLP}_{h^*}\right) \qquad \forall h^* \qquad (7.287)$$

$$\lambda_{h^*} = \sum_r x^{CLP}_{h^*,r} \lambda_r \qquad \forall h^* \qquad (7.288)$$

$$\lambda_{h^*+1} = \sum_r x^{CLP}_{h^*+1,r} \lambda_r \qquad \forall h^* \qquad (7.289)$$

$$\lambda_{h^*}(t) = \lambda_{h^*} + \left( \frac{\lambda_{h^*+1} - \lambda_{h^*}}{\mathrm{t_{h^*+1}} - \mathrm{t_{h^*}}} \right) \left( t - \mathrm{t_{h^*}} \right) \qquad \forall h^* \qquad (7.290)$$

$$ICF_{i,h^*}(t) = \mathrm{ICF_{i,h^*}} + \left( \frac{\mathrm{ICF_{i,h^*+1}} - \mathrm{ICF_{i,h^*}}}{\mathrm{t_{h^*+1}} - \mathrm{t_{h^*}}} \right) \left( t - \mathrm{t_{h^*}} \right) \qquad \forall i \in INT, h^* \qquad (7.291)$$

$$REV_{i,h^*}(t) = \left( \lambda_{h^*} + \left( \frac{\lambda_{h^*+1} - \lambda_{h^*}}{\mathrm{t_{h^*+1}} - \mathrm{t_h}} \right) \left( t - \mathrm{t_{h^*}} \right) \right) \left( \mathrm{ICF_{i,h^*}} + \left( \frac{\mathrm{ICF_{i,h^*+1}} - \mathrm{ICF_{i,h^*}}}{\mathrm{t_{h^*+1}} - \mathrm{t_h}} \right) \left( t - \mathrm{t_{h^*}} \right) \right) \qquad \forall i \in INT, h^* \qquad (7.292)$$

$$AvgREV_{i,h^*} = \frac{1}{t_{h^*+1} - t_h} \int_{t_{h^*}}^{t_{h^*+1}} REV_{i,h^*}(t)\, dt \qquad \forall i \in INT, h^* \qquad (7.293)$$

# REFERENCES


AEMO. (2010). *Introduction to Australia's National Electricity market* (pp. 1–28).

Alessandri, T. M., Ford, D. N., Lander, D. M., Leggio, K. B., & Taylor, M. (2004). Managing risk and uncertainty in complex capital projects. *The Quarterly Review of Economics and Finance*, *44*(5), 751–767. http://doi.org/10.1016/j.qref.2004.05.010

Arrow, K. J., & Debreu, G. (1954). Existence of an Equilibrium for a Competitive Economy. *Econometrica: Journal of the Econometric Society*, *22*(3), 265–290.

Artzner, P., Delbaen, F., Eber, J. M., & heath, D. (1998). Coherent Measures of Risk.

Baldick, R. (2009). Augmented Screening Curve Approach to Optimizing Generation Capacity Additions Considering Reserves (pp. 1–26).

Bank, C. (1992). The Chase Guide to Risk Management Products. Chase Bank.

Barroso, L. A., Granville, S., & Trinkenreich, J. (2003). Managing hydrological risks in hydro-based portfolios. ... *General Meeting*, *2*. http://doi.org/10.1109/PES.2003.1270395

Batstone, S. R. (2000). An equilibrium model of an imperfect electricity market. *Department of Management, University of Canterbury, New Zealand*.

Batstone, S. R. (2003). Aspects of risk management in deregulated electricity markets: Storage, market power and long-term contracts. *Department of Management, University of Canterbury, New Zealand*, *Ph.D*.

Baumol, W. J. (1968). On the social rate of discount.

Bean, J. C., Higle, J. L., & Smith, R. L. (1992). Capacity expansion under stochastic demands. *Operations Research*, *40*, 210–216.

Bell, G. (2010). Living in a Carbon-based World: CO2 and its impact on the EU Power Sector.

Bernard, J.-T., & Chatel, J. (1984). The role of energy limited hydro equipment in an optimal plant mix. *Energy Economics*, *6*(2), 139–144. http://doi.org/10.1016/0140-9883(84)90029-X

Bjorndal, M., & Jornsten, K. (2008). Equilibrium prices supported by dual price functions in markets with non-convexities. *European Journal of Operational Research*, *190*(3), 768–789.

Blavatskyy, P. (2010). Modifying the Mean-Variance Approach to Avoid Violations of Stochastic Dominance. *Management Science*.

Blyth, W. (2007). *Climate Policy & Investment Risk* (pp. 1–144).

Bohn, R. E., Caramanis, M. C., & Schweppe, F. C. (1984). Optimal Pricing in Electrical Networks over Space and Time. *The RAND Journal of Economics*, 360–376.

Borenstein, S., Bushnell, J. B., & Stoft, S. (2000). The competitive effects of transmission capacity in a deregulated electricity industry. *The RAND Journal of Economics*, *31*(2), 294–325.

Botterud, A., & Doorman, G. L. (2008). Generation Investment and Capacity Adequacy in Electricity Markets. *International Association for Energy Economics. ...*, 11–15.

Botterud, A., & Korpås, M. (2007). A Stochastic Dynamic Model for Optimal Timing of Investments in New Generation Capacity in Restructured Power Systems. *International Journal of Electrical Power \& Energy Systems*, *29*(2), 163–174. http://doi.org/10.1016/j.ijepes.2006.06.006

Botterud, A., Ilic, M., & Wangensteen, I. (2005). Optimal investments in power generation under centralized and decentralized decision making. *IEEE Transactions on Power Systems*, *20*(1), 254–263.

Boucher, J., & Smeers, Y. (2012). Energy security and long-term arrangements, 1–25.

Bunn, D. W. (1984). Applied Decision Analysis.

Bushnell, J. B. (2003). A mixed complementarity model of hydrothermal electricity competition in the western United States. *Operations Research*, 80–93.

Caramanis, M. C. (1982). Investment decisions and long-term planning under electricity spot pricing. *IEEE Transactions on Power Apparatus and Systems*, 4640–4648.

Caramanis, M. C., Bohn, R. E., & Schweppe, F. C. (1982). Optimal spot pricing: practice and theory. *IEEE Transactions on Power Apparatus and Systems*, 3234–3245.

Charnes, A., Cooper, W. W., & Symonds, G. H. (1958). Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management Science*, *4*(3), 235–263. http://doi.org/10.1287/mnsc.4.3.235

Conejo, A. J., Carrión, M., & Morales, J. M. (2010). Decision Making Under Uncertainty in Electricity Markets (Vol. 153). Boston, MA: Springer US. http://doi.org/10.1007/978-1-4419-7421-1

Cramton, P. C. (2010). Using Forward Markets to Improve Electricity Market Design. *Utilities Policy*.

Cramton, P. C., & Stoft, S. (2005). A capacity market that makes sense. *The Electricity Journal*, *18*(7), 43–54.

Deng, S. J., & Oren, S. S. (2006). Electricity derivatives and risk management. *Energy*, *31*(6-7), 940–953–953.

Dentcheva, D., & Ruszczyński, A. P. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking \& Finance*, *30*(2), 433–451. http://doi.org/10.1016/j.jbankfin.2005.04.024
Dixit, A. K. (2012). Investment under uncertainty.
Doege, J., Schiltknecht, P., & Lüthi, H. J. (2006). Risk management of power portfolios and valuation of flexibility. *Or Spectrum*, *28*(2), 267–287–287.
Dow, J., & Ribeiro da Costa Werlang, S. (2003). Uncertainty Aversion, Risk Aversion, and the Optimal Choice Portfolio. *Econometrica*, *60*(1), 197–204.
Dye, S. (1994). *On a Flexible Model for New Zealand's Hydro-Thermal Electricity Generation System*. mro.massey.ac.nz.
Ehrenmann, A., & Smeers, Y. (2011). Generation Capacity Expansion in a Risky Environment: A Stochastic Equilibrium Analysis. *Operations Research*, *59*(6), 1332–1346.
Evans, L. T., & Guthrie, G. A. (2005). Risk, price regulation, and irreversible investment. *International Journal of Industrial Organization*, *23*(1-2), 109–128. http://doi.org/doi: DOI: 10.1016/j.ijindorg.2004.11.005
Eydeland, A., & Geman, H. (1999). Fundamentals of electricity derivatives. *Energy Modelling and the Management of …*.
Fabian, C. I., & Veszpremi, A. (2008). Algorithms for handling CVaR constraints in dynamic stochastic programming models with applications to finance. *Journal of Risk*, *10*(3), 111.
Farrar, D. (1964). The Investment Decision Under Uncertainty. The Ford Foundation Dissertation Series.
Finon, D. (2008). Investment risk allocation in decentralised electricity markets. The need of long-term contracts and vertical integration.
Fishburn, P. C. (1984). Foundations of risk measurement. I. Risk as probable loss. *Management Science*.
Francis, J., & Archer, S. (1979). Portfolio Analysis. Foundation of Finance Series.
Garces, L., Conejo, A. J., García-Bertrand, R., & Romero, R. (2009). A bilevel approach to transmission expansion planning within a market environment. *IEEE Transactions on Power Systems*, *24*(3), 1513–1522.
Genc, T. S., Reynolds, S. S., & Sen, S. (2007). Dynamic oligopolistic games under uncertainty: A stochastic programming approach. *Journal of Economic Dynamics and Control*, *31*(1), 55–80.
Geoffrion, A. (1976). The purpose of mathematical programming is insight, not numbers. *Interfaces*, 25.
Green, R. (2002). Competition in generation: The economic foundations. *Proceedings of the IEEE*, *88*(2), 128–139–139.
Guzelsoy, M. (2007). Duality for mixed-integer linear programs. *Internat J Oper Res*.
Harker, P. T., & Pang, J.-S. (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming*, *48*(1), 161–220.
Hobbs, B. F., Metzler, C., & Pang, J.-S. (2000). Strategic gaming analysis for electric power systems: an MPECapproach. *IEEE Transactions on Power Systems*, *15*(2), 638–645.
Hogan, W. W., Read, E. G., & Ring, B. J. (1996). Using mathematical programming for electricity spot pricing. *International Transactions in Operational Research*, *3*(3-4), 209–221.
Huang, Y.-H., & Wu, J.-H. (2008). A portfolio risk analysis on electricity supply planning. *Energy Policy*, *36*(2), 627–641. http://doi.org/10.1016/j.enpol.2007.10.004
Jagannathan, R., & McGrattan, E. (1995). The CAPM Debate. *Federal Reserve Bank of Minneapolis Quaterly Review*.
Joskow, P. L. (2000). Transmission rights and market power on electric power networks. *The RAND Journal of Economics*.
Kallberg, J., & Ziemba, W. T. (1983). Comparison of Alternative Utility Functions in Portfolio Selection Problems. *Management Science*.
Kettunen, J. (2008). Electricity Investment Behavior in Response to Climate Policy Risk. *Iaee.org*.
Knight, F. (1921). Risk, Uncertainty & Profit.
Krokhmal, P., & Uryasev, S. P. (2003). Numerical comparison of CVaR and CDaR approaches: Application to hedge funds. *Applications of Stochastic …*.
Krokhmal, P., Palmquist, J., & Uryasev, S. P. (2002). Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints. *Journal of Risk*.
Ku, A. (1995). *Risk And Flexibility In Electricity Markets*.
Kunzi-Bay, A., & Mayer, J. (2006). Computational aspects of minimizing conditional value-at-risk. *Computational Management Science*, *3*(1), 3–27. http://doi.org/10.1007/s10287-005-0042-0

Layton, B. (2007). The markets for electricity in New Zealand, 1–73.
Lee, Y., & Oren, S. S. (2009). An equilibrium pricing model for weather derivatives in a multi-commodity setting. *Energy Economics*.
Levin, N., Tishler, A., & Zahavi, J. (1985). Capacity expansion of power generation systems with uncertainty in the prices of primary energy resources. *Management Science*, 175–187.
Leyffer, S. (2009). A Complementarity Constraint Formulation of Convex Multiobjective Optimization Problems. *INFORMS Journal on Computing*, *21*(2), 257–267–267. http://doi.org/10.1287/ijoc.1080.0290
Lino, P., Barroso, L. A., Pereira, M. V., Kelman, R., & Fampa, M. H. C. (2003). Bid-based dispatch of hydrothermal systems in competitive markets. *Annals of Operations Research*, *120*(1), 81–97.
Liu, M., Wu, F. F., & Ni, Y. (2006). A survey on risk management in electricity markets (p. 6). Presented at the Power Engineering Society General Meeting, 2006. IEEE.
Macgill, I. (2010). Electricity market design for facilitating the integration of wind energy: Experience and prospects with the Australian National Electricity Market. *Energy Policy*, *38*(7), 3180–3191. http://doi.org/10.1016/j.enpol.2009.07.047
Majumdar, S., & Chattopadhyay, D. (1999). A model for integrated analysis of generation capacity expansion and financial planning. *IEEE Transactions on Power Systems*, *14*(2), 466–471.
Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 77–91.
Markowitz, H. M. (1991). Foundations of portfolio theory. *Journal of Finance*, *46*(2), 469–477.
Meade, R., & O'Connor, S. (2009). Comparison of long-term contracts and vertical integration in decentralised electricity markets.
Mitropoulos, C. (1984). Determining the optimal reserve capacity margin in electricity supply: a note. *Journal of the Operational Research Society*, 647–652.
Murphy, F. H., & Mudrageda, M. V. (1998). A decomposition approach for a class of economic equilibrium models. *Operations Research*, *46*(3), 368–377.
Murphy, F. H., & Smeers, Y. (2005). Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations Research-Baltimore Then Linthicum-*, *53*(4), 646.
Neuhoff, K., & De Vries, L. (2004). Insufficient incentives for investment in electricity generations. *Utilities Policy*, *12*(4), 253–267–267.
Noyan, N. (2010). *Optimization With First Order Stochastic Dominance Constraints*.
NZEC. (2007). *Electricity Commission SOO Scenario Analysis - Discount Rates*.
Ogryczak, W., & Ruszczyński, A. P. (2002). Dual Stochastic Dominance and Quantile Risk Measures. *International Transactions in Operational Research*, *9*(5), 661–680.
Oren, S. S. (2005). Ensuring generation adequacy in competitive electricity markets. *Electricity Deregulation: Choices and Challenges*.
Pang, J.-S. (2004). On the global minimization of the value-at-risk. *Optimization Methods and Software*.
Pereira, M. V., & Campodónico, N. (1999). Application of stochastic dual DP and extensions to hydrothermal scheduling. Online Rep.
Pflug, G. C. (2000). Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. In *Probabilistic Constrained Optimization* (Vol. 49, pp. 272–281). Boston, MA: Springer US. http://doi.org/10.1007/978-1-4757-3150-7_15
Prékopa, A. (1973). Contributions to the theory of stochastic programming. *Mathematical Programming*, *4*(1), 202–221. http://doi.org/10.1007/BF01584661
Ralph, D., & Smeers, Y. (2006). EPECs as models for electricity markets (pp. 74–80). Presented at the Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES. http://doi.org/10.1109/PSCE.2006.296252
Ralph, D., & Smeers, Y. (2011). Pricing risk under risk measures:an introduction to stochastic-endogenous equilibria, 1–45.
Ralph, D., & Smeers, Y. (2015). Risk Trading and Endogenous Probabilities in Investment Equilibria. SIAM Journal on Optimization, 2015, Vol 25, No4, pp2589-2611
Read, E. G., & Hindsberger, M. (2010). Constructive dual DP for reservoir optimization. *Handbook of Power Systems I*, 3–32.
Read, E. G., & Thomas, M. (2005). Risk-Adjusted Discount Rates and Optimal Plant Mix: A Conceptual Analysis for Electricity Markets.
Robichek, A. A., & Myers, S. C. (1966). Conceptual problems in the use of risk-adjusted discount rates.
Rockafellar, R. T., & Uryasev, S. P. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*, 21–42–42.
Roques, F. A., & Newbery, D. M. G. (2008). Fuel mix diversification incentives in liberalized

electricity markets: A Mean-Variance Portfolio theory approach. *Energy Economics*. Retrieved from http://ac.els-cdn.com.ezproxy.canterbury.ac.nz/S0140988307001478/1-s2.0-S0140988307001478-main.pdf?_tid=e5b8bbd0dab7037b7dd783a5a2a89c01&acdnat=1345866142_bd9aeb25f9171511893193aa72fc8d60

Rothkopf, M. H., O Neill, R. P., Hobbs, B. F., Sotkiewicz, P. M., & Stewart, W. R., Jr. (2004). Price Tests for Entry into Markets in the Presence of Non-Convexities.

Sarykalin, S., & Serraino, G. (2008). Value-at-Risk vs. Conditional Value-at-risk in risk management and optimization. *2008 Tutorials in …*.

Shanbhag, U. V., Infanger, G., & Glynn, P. W. (2011). A Complementarity Framework for Forward Contracting Under Uncertainty. *Operations Research*, *59*(4), 810–834.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.

Sherali, H., Soyster, A., Murphy, F. H., & Sen, S. (1982). Linear Programming Based Analysis of Marginal Cost pricing in Electric Utility capacity Expansion. *European Journal of Operational Research*.

Sriboonchitta, S. (2010). Stochastic Dominance and Applications to Finance, Risk and Economics (pp. 1–442).

Stewart, P. (2007, March 15). *Intertemporal Considerations In Supply Offer Development In THe Wholesale Electricity Market*.

Stoft, S. (2002). Power System Economics. Wiley-IEEE Press.

Stridbaek, U. (2005). Lessons From Liberalised Electricity Markets (pp. 1–223).

Traber, T., & Kemfert, C. (2011). Gone with the wind? — Electricity market prices and incentives to invest in thermal power plants under increasing wind energy supply. *Energy Economics*, *33*(2), 249–256. http://doi.org/10.1016/j.eneco.2010.07.002

Wallace, S. W. (2009). Delta-hedging a hydropower plant using stochastic programming. *Optimization in the Energy Industry*.

Wallace, S. W. (2010). Stochastic programming and the option of doing it differently. *Annals of Operations Research*, *177*(1), 3–8.

Weber, C. (2011). Uncertainty in the Electric Power Industry (pp. 1–312).

Wu, A. (2012). Modelling Generation Investment under Wind Induced Uncertainty, 1–29.

Yang, M., & Read, E. G. (1999). A constructive dual dynamic programing for a reservoir model with correlation. *Water Resources Research*, *35*(7), 2247–2257.

Yao, J., Adler, I., & Oren, S. S. (2008). Modeling and Computing Two-Settlement Oligopolistic Equilibrium in a Congested Electricity Network. *Operations Research*, *56*(1).

Zhao, Y., & Ziemba, W. T. (2001). A stochastic programming model using an endogenously determined worst case risk measure for dynamic asset allocation. *Mathematical Programming*, *89*(2), 293–309.